

Seminar

Einführung ins wissenschaftliche

Arbeiten- Clustering

WiSe18/19

16.11.2018

Übersicht

- **17.10.18, 14:00 Uhr:**
 - Seminarvorstellung
 - Zuteilung Themen Phase 1
- **16.11.18, 10:00- 18:00, Raum 169 :**
 - Vorträge Phase 1
 - Jeweils 10-15 Minuten + 5 Minuten Fragen
 - Zuteilung Themen Phase 2
- **21.12.18, 12:00 Uhr mittags:**
 - Abgabe der schriftlichen Ausarbeitungen zu Phase 2
- **18.01.19 und 25.01.19, 10:00- 16:00, Raum C007:**
 - Vorträge Phase 2
 - Jeweils 20-25 Minuten + 10 Minuten Fragen

Hinweise

- Alle Themen werden verlost
- Es gibt Templates für Folien und Ausarbeitung:
http://www.dbs.ifi.lmu.de/cms/studium_lehre/bama_theses/index.html#BA
- **Alle Vortragsfolien per Mail am Tag vorher bis 16:00 Uhr schicken.**
Format: PDF oder PowerPoint.
Benennung: Vortrag<Nummer des Themas><Thema>.pdf
(Beispiel: Vortrag01IntroductionDataMiningandClustering.pdf)
- Die **Ausarbeitung** wird mit LaTeX gemacht, Abgabe als PDF bis spätestens **21.12.18, 12 Uhr mittags, über uniworx** (5 Wochen Zeit).
Benennung: Ausarbeitung<Nummer des Themas><Thema>.pdf
- Sprecht euch mit den Vortragenden ab, die euer Partnerthema haben (Themennummer mod 8 ist gleich)

Phase 2: Ausarbeitungen- Formal

- Sprache: Englisch (empfohlen) / deutsch
- Abgabe bis 21.12.18, 12 Uhr **mittags** über uniworx
- Umfang: 7 000 – 14 000 Zeichen
- Freiwilliges (aber empfehlenswertes) Zwischenstandstreffen mit mir
- Auf das achten, was ihr gerade gelernt habt, u.a.:
 - Korrektes zitieren, alle Quellen angeben, nicht einfach übersetzen/abschreiben
 - Wissenschaftlicher Schreibstil
 - Ansprechende, ordentliche Form. Gerne eigene Abbildungen erstellen- diese haben im Paper häufig einfach nur keinen Platz
 - Eigene Recherche zum Thema durchführen
 - Aufbau und Sprache klar, einfach zu verstehen, und fehlerfrei
 - **Kritische** Auseinandersetzung mit den Papern. Die Seminararbeit soll nicht versuchen, das Thema zu „verkaufen“- stattdessen sowohl Vor- als auch Nachteile herausarbeiten!
 - ...

Phase 2: Überblick

Tag 1: Einführung in verschiedene Gebiete des Clustering

- Überblick geben
- Grundlagen und bekannteste Ansätze erklären
- Herausfinden, was für das Gebiet wichtig ist
- In das Thema einführen
- Welche Problemstellungen gibt es?
- ...

Phase 2: Überblick

Tag 2: Auszug aktueller Forschung in diesen Gebieten

- (Nicht nur) das jeweilige Paper in der Tiefe verstehen
- Wichtige Related Work dazu lesen und erläutern
- Was ist neu/ anders an dem Ansatz?
- Eigenes Anwendungsbeispiel als roten Faden verwenden
- Bei 2 Shortpapern: Unterschiede und Ähnlichkeiten
- Auch hier muss kurz in das Thema eingeführt werden
- Was ist die Problemstellung die gelöst wird?
- ...

Phase 2: Themen Tag 1

1. Introduction: Data Mining and Clustering
2. Introduction: Subspace and Correlation Clustering
3. Introduction: Clustering of Time Series, Process Mining and Streams
4. Introduction: Interactive and Visual Clustering
5. Introduction: Clustering of Big Data, Distributed Clustering
6. Introduction: Graphs and Networkanalysis
7. Introduction: Anytime Clustering
8. Introduction: Outlier Detection

Phase 2: Themen Tag 2 i)

9. Specialization: Density-based Clustering
 - Study on density peaks clustering based on k-nearest neighbors and principal component analysis (Journal)
 - Clustering by fast search and find of density peaks (Journal)
10. Specialization: Subspace Clustering
 - Discovering Non-Redundant K-means Clusterings in Optimal Subspaces
11. Specialization: Clustering of Time Series
 - k-Shape: Efficient and Accurate Clustering of Time Series
12. Specialization: Interactive Clustering
 - PARADISO: An Interactive Approach of Parameter Selection for the Mean Shift Algorithm
 - Interactive Time Series Clustering with COBRAS

Phase 2: Themen Tag 2 ii)

13. Specialization: Clustering in Large Graphs
 - Network Lasso: Clustering and optimization in Large Graphs
14. Specialization: Clustering in Graphs
 - Local Higher-Order Graph Clustering
15. Specialization: Anytime Clustering
 - AnyDBC: An Efficient Anytime Density-based Clustering Algorithm for Very Large Complex Datasets
16. Specialization: Outlier Detection
 - Reverse Nearest Neighbors in Unsupervised Distance-Based Outlier Detection

Verlosung



Fragen?

