

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

Seminar: Ausgewählte Themen der Wissensverarbeitung Vorstellung der Themen

Prof. Dr. Stumme, Dr. Hanika

2019-10-21

Wintersemester 2019/2020

Allgemeines

Inhalt

Ausgewählte Themen (KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

Das Seminar behandelt verschiedene Methoden

- zur Wissensentdeckung in Daten,
- zur Wissensrepräsentation
- insbesondere zur Analyse von (sozialen) Netzwerken.

Hierbei betrachten wir u.a.:

- überwachte / unüberwachte Lernverfahren
- Metriken in Daten, insbesondere Netzwerken
- Anomalieerkennung
- Charakterisierung von sozialen Netzwerken
- Approximation und Exploration von Wissen
- Ordnungsstrukturen in Wissen
- Ontologieentwicklung.

Für Wen? Was gibt es?

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

Der angesprochene Hörer_innenkreis ist:

- Informatik Bachelor, ab 4. Semester
- Informatik Master
- Mathematik Bachelor/Master mit NF Informatik

Vorkenntnisse:

- Informatik Grundstudium für Einbringung in den Bachelor.
- Abgeschlossener Bachelor für Einbringung in den Master.
- Freude wissenschaftliche Artikel zu lesen!

Leistungsnachweis / Umfang:

- Hausarbeit (in der Form eines wiss. Artikels)
- Referat / Präsentation
- Umfang ist 2 SWS

Ablauf: Themenausgabe und Ersttreffen

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

- Zu Beginn wird pro Thema jeweils ein Artikel ausgegeben,
- der durch den / die Seminarteilnehmer_in im Laufe des Semesters durch weitergehende Literatur ergänzt wird.
 - Die Auswahl der weiteren Literatur trifft der / die Seminarteilnehme_in in Absprache mit dem Betreuer.
 - Eine Abmeldung von einem Thema ist **maximal eine Woche** nach Ausgabe des Themas möglich, um dann anderen Studierenden die Möglichkeit der Teilnahme zu geben.
 - Idealerweise trifft man sich circa nach **4-6 Wochen** mit der betreuenden Person und stellt die weiteren Quellen vor. Die Auswahl geht in die Endnote ein.

Ziel: ein eigener Überblicksartikel zum gewählten Thema.

Ablauf: Abgabe und Vortrag

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

Abgabetermin:

Die Arbeit ist am Montag, 3. Februar 2020, abzugeben.

Die Arbeit

- muss in $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ erstellt werden.
- muss im Springer Style (LNCS) sein.
- hat einen Umfang von 12 (16) Seiten zu haben für Bachelor (Master).

Zum Vortrag:

- Die Vorträge finden in der Woche 10-14.02.2020 statt.
- Drei (3!) Tage vor dem geplanten Vortrag sind per Email als PDF abzugeben: a) Eine einseitige Zusammenfassung; b) der Foliensatz

Themen

Themen

Betreut durch Maximilian Stubbemann

Efficient community detection using power graph analysis [6]

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

Auszug aus dem Abstract des Artikels

[...] we apply the Power Graph Analysis methodology, for the first time to the best of our knowledge, to the field of community mining. The model was introduced in bioinformatics research and in this work is applied to the problem of community detection in complex networks. The advances in the field of community mining allow us to experiment with widely accepted benchmark data sets [...]

TODO:

Stellen Sie die Grundlagen der Power Graph Analysis vor und beschreiben Sie wie diese im vorliegenden Paper für die Community Detection genutzt wird. Sammeln Sie weitere Literatur zum Thema Power Graph Analysis um weitere Anwendungsfelder zu finden.

Paper2vec: Combining Graph and TextInformation for Sci. Papers [7]

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

Auszug aus dem Abstract des Artikels

[...] a novel neural network embedding based approach for creating scientific paper representations which make use of both textual and graph-based information. [...] We demonstrate the efficacy[...] in two tasks - node classification and link prediction[...]

TODO:

Beschreiben Sie wie das vorliegende Paper Dokumenteneinbettungsverfahren mit Knoteneinbettungsmethoden kombiniert um wissenschaftliche Publikation durch niedrigdimensionale Vektoren darzustellen. Die Vorstellung der genutzten Techniken für Dokumenteinbettungen (doc2Vec) und Knoteneinbettungen (Deepwalk) ist dabei ein wichtiger Teil der Seminararbeit.

Themen

Dominik Dürschnabel

Linear Time Transitive Orientation of Comparability Graphs [1]

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

**Dominik
Dürschnabel**

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

Auszug aus dem Abstract des Artikels

[...] There is a canonical linear-space representation for the modules of a graph, called the modular decomposition.[...] We give the first linear-time algorithm for modular decomposition, and a new bound of on transitive orientation [...]

TODO:

Die vorliegende Arbeit beschreibt wie Graphen einer bestimmten Graphenklasse in Linearzeit in Module zerlegt werden können. Diese Module können dazu genutzt werden, diesen Graphen ebenfalls in Linearzeit transitiv zu orientieren. Erklären und beweisen Sie den daraus entstehenden Algorithmus mithilfe von weiteren, selbst ausgewählten Veröffentlichungen.

Themen

Betreut durch Maximilian Felde

Finding the Number of Clusters in a Dataset [2]

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

Auszug aus dem Abstract des Artikels

[...] In this article we develop a simple, yet powerful nonparametric method for choosing the number of clusters based on distortion, a quantity that measures the averaged distance, per dimension, between each observation and its closest cluster center. [...]

TODO:

Die vorliegende Arbeit beschreibt einen informationstheoretischen Ansatz zur Identifikation der Anzahl Cluster in einem Datensatz. Fassen Sie die Ergebnisse dieser Veröffentlichung zusammen und vergleichen Sie diesen Ansatz mit anderen Ihnen bekannten Ansätzen zum Bestimmen der Clusteranzahl. Ziehen Sie dazu weitere Literatur heran.

The Random Subspace Method [3]

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

Auszug aus dem Abstract des Artikels

[...] The classifier consists of multiple trees constructed systematically by pseudorandomly selecting subsets of components of the feature vector, that is, trees constructed in randomly chosen subspaces. [...]

TODO:

Die vorliegende Arbeit beschreibt die Random Subspace Methode zur Verbesserung der Klassifikationsleistung von Entscheidungsbäumen. Fassen Sie die Ergebnisse dieser Veröffentlichung und 2-3 weiterer verwandter Arbeiten zusammen. Sie können hierbei beispielsweise auf die Verbreitung der Random Subspace Methode eingehen oder die Methode mit anderen Ensemble Learning Methoden vergleichen.

Themen

Betreut durch Bastian Schäfermeier

On the predictability of talk attendance at academic conferences [4]

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

Auszug aus dem Abstract des Artikels

[...] This paper focuses on the prediction of real-world talk attendances at academic conferences with respect to different influence factors.[...] We apply Hybrid Rooted Page-Rank, a state-of-the-art unsupervised machine learning method that combines information from different sources. [...]

TODO:

Die vorliegende Arbeit beschreibt Möglichkeiten, wie der Besuch von Vorträgen einer Konferenz vorhergesagt werden kann. Die Vorhersagen basieren dabei auf Informationen aus sozialen Netzwerken sowie aus den Forschungsthemen der Konferenzteilnehmer. Fassen Sie die Ergebnisse dieser Veröffentlichung sowie die Inhalte von 2 selbst ausgewählten verwandten Arbeiten zusammen. Sie können dabei selbst entscheiden, auf welchen Aspekt Sie sich konzentrieren.

Learning the parts of objects by non-negative matrix factorization [5]

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

Auszug aus dem Abstract des Artikels

[...] we demonstrate an algorithm for non-negative matrix factorization that is able to learn parts of faces and semantic features of text. [...] When non-negative matrix factorization is implemented as a neural network, parts-based representations emerge by virtue of two properties: the firing rates of neurons are never negative and synaptic strengths do not change sign.[...]

TODO:

Die vorliegende Arbeit beschreibt den Algorithmus NMF zur Faktorisierung von Matrizen. Erklären Sie die in der Arbeit vorgestellte Methodik und erläutern Sie die Unterschiede zu anderen Methoden der Matrixfaktorisierung. Suchen Sie 2-3 weitere Arbeiten, die Erweiterungen oder Varianten von NMF darstellen und fassen Sie diese zusammen.

Themen

Betreut durch Maren Koyda

Reducing the Representation Complexity of Lattice-Based Taxonomies [8]

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

Auszug aus dem Abstract des Artikels

[...] We propose a method where the notion of stability is introduced to select potentially more pertinent concepts. We present some theoretical properties of stability and discuss several use cases where taxonomy building is an issue [...]

TODO:

Die vorliegende Arbeit beschreibt die Stabilität als Maß auf formalen Begriffen. Erläutern Sie dieses Maß sowie seine Anwendungsmöglichkeiten und vergleichen Sie es mit anderen Maßen auf formalen Begriffen aus weiterführender Literatur.

Themen

Betreut durch Tom Hanika

Scale-free networks are rare [9]

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

Auszug aus dem Abstract des Artikels

Real-world networks are often claimed to be scale free [...] we find robust evidence that strongly scale-free structure is empirically rare, while for most networks, log-normal distributions fit the data as well or better than power laws.[...]

TODO:

Das scale-free network-Modell (SFN) hat sich seit 1999 weit verbreitet und viel Anwendung gefunden. Die vorliegenden Schriften allerdings widerlegen genau diese breite Anwendbarkeit. Ausgehend von den beiden Artikeln, recherchieren Sie nach weiteren Artikeln pro und contra zum SFN. Erstellen Sie einen Übersichtsartikel zum SFN und den aktuellen Stand der Debatte. Fokussieren Sie sich dabei insbesondere auf einen von Ihnen gewählten Aspekt / Eigenschaft / Anwendung des Modells.

Many-Valued Conceptual Data [10]

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

Auszug aus dem Abstract des Artikels

[...] Conceptual scaling has been developed in the frame of formal concept analysis, a theory based on a mathematization of conceptual hierarchies [...] these scalings are used to introduce and to study a general notion of dependency between attributes which covers special notions like functional and linear dependency[...]

TODO:

Erstellen Sie einen Übersichtsartikel über die in *Conceptual Scaling* vorgestellte Methode und finden Sie Erweiterungen der Methoden. Gehen Sie dabei insbesondere auf die verschiedenen Notwendigkeiten ein die sich aus der Datenart ergeben als auch die Auswirkungen auf funktionale Abhängigkeiten.

Bounds in Information Theory [11]

Auszug aus dem Abstract des Artikels

[...] purpose of this paper is to present a new bound for the Shannon Entropy [...]

TODO:

Das Konzept der Entropy wird in der Informatik an vielen Stellen genutzt. Insbesondere findet es auch Anwendung in der Wissensverarbeitung bei überwachten und unüberwachten Lernverfahren. Erstellen Sie einen Übersichtsartikel über häufig verwendete Entropy-Maße mit Fokus auf Graph-Daten. Erklären Sie inwiefern die in der vorliegenden Arbeit vorgestellte Methode zur Abschätzung der Shannon-Entropy hilfreich ist. Stellen Sie darüber hinaus die Anwendung zur Messung von struktureller Komplexität in sozialen Netzwerken vor und vergleichen Sie diese mit anderen Komplexitätsmaßen.

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

Robust Rules in Data [12]

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

Auszug aus dem Abstract des Artikels

[...] we identify the *best set* of rules as the one that most succinctly describes the data. We show that the resulting optimization problem does not lend itself for exact search, and hence propose Grab, a greedy heuristic [...]

TODO:

Regeln in Daten finden ist eine wichtige Anwendung in der Wissensverarbeitung. Die vorliegende Arbeit stellt ein neues Verfahren vor um eine Menge von 'wichtigen' Regeln zu identifizieren. Vergleichen Sie das Vorgehen mit bekannten Methoden zu Assoziationsregeln. Finden Sie weitere alternative Methoden. Erstellen Sie dazu einen Übersichtsartikel bei welchem Sie nach selbstgewählten Kriterien die verschiedenen Methoden vergleichen.

Rule Mining in Knowledge Graphs [13]

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

Auszug aus dem Abstract des Artikels

[...] Our approach AMIE (Galárraga et al. in WWW, 2013) has shown how rules can be mined effectively from KBs even in the absence of counterexamples. In this paper, we show how this approach can be optimized to mine even larger KBs with more than 12M statements.[...]

TODO:

Rule-Mining in Knowledge Graphs ist aufgrund ihrer sehr komplexen Struktur sehr aufwendig und komplex. Erstellen Sie ausgehend von dem genannten Artikel eine Übersichtsarbeit zu Rule-Mining in Knowledge Graphs. Strukturieren Sie hierfür Ihre Arbeit insbesondere hinsichtlich der unterschiedlichen Ausdrucksstärke der verschiedenen Arten von Rules.

Projekte

Auflistung der Projekte

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

Legende: M = methodischer Schwerpunkt, T = technischer Schwerpunkt, B = als Bachelorarbeit möglich, MA = als Masterarbeit möglich

- PCA auf formalen Kontexten T,M
- Begriffliches Skalieren von Datensätzen T,M
- Begriffsverband der Maße T,M,B
- Begriffliche Datenanalyse auf großen Datensätzen T,M,B
- Standard-Sample-Set für die Algorithmenentwicklung sozialer Netzwerke T,M
- BibSonomy: Weiterentwicklung des BibTeX-Parsers T,M
- BibSonomy: Personen- und Autoren-Clouds T,M
- BibSonomy: Neustrukturierung der Community Post Seite T,M

Auflistung der Projekte

Ausgewählte
Themen
(KDE)

Stumme,
Hanika

Allgemeines

Themen

Betreut durch
Maximilian
Stubbemann

Dominik
Dürschnabel

Betreut durch
Maximilian Felde

Betreut durch
Bastian Schäfermeier

Betreut durch Maren
Koyda

Betreut durch Tom
Hanika

Projekte

- Dichtebasiertes Clustering und FBA T,M,B
- Entropie in formalen Kontexten T,M,B,MA
- Hubs and Authorities in Formal Contexts T,M,BA
- Kernel Systems vs Closure Systems T,M,BA
- Series Parallel Networks in Citation Networks T,M,BA
- Generator für zufällige formale Kontexte T,M
- Pedestrian Dead Reckoning Methoden zur Schätzung der Lage von Orten zueinander T,M,BA,MA
- Framework zur Annotation und Klassifikation von Zeitreihen T,M