

Seminar Wissensbasierte Systeme

Wintersemester 2013 / 2014

Prof. Dr. Hans Kleine Büning

Dr. Maik Anderka

24.10.2013

Themenbereiche

- Im Seminar werden dieses Semester Themen aus den Bereichen **maschinelles Lernen** und **Data-Mining** behandelt
- Konkrete Themen könnten aus den folgenden Teilbereichen stammen:
 - Data stream mining
 - Anomaly detection, outlier analysis
 - Semi-supervised learning
 - Web-Mining, Text-Mining
 - Big data analytics
 - ...

Auswahl von Themen

- Eine Auswahl von Themenvorschlägen wird bereitgestellt
- Eigene Themenvorschläge nehmen wir gern entgegen
- Themen sollten behandeln:
interessante **theoretische** oder **anwendungsorientierte**
Fragestellungen aus den genannten Bereichen
- Themen müssen in **Breite und Tiefe** bearbeitet werden.
Dazu gehört:
 - ein vergleichender Überblick über existierende Verfahren zur Lösung der Fragestellung sowie
 - eine detaillierte Beschreibung mindestens eines dieser Verfahren, inklusive Theorie, Algorithmus und erzielten Ergebnissen

Prüfungsleistungen

1. termingerechte Abgabe einer sinnvollen **Gliederung**
2. termingerechte Abgabe der schriftlichen **Ausarbeitung**
3. termingerechte Abgabe der **Peer-Reviews** zu 2 anderen Ausarbeitungen
4. Während der Blocktermine **Vortrag** von 35-45 Minuten Dauer anhand von Folien
5. während der Blocktermine **Anwesenheit und Beteiligung** an den Diskussionen
6. termingerechte Abgabe der entsprechend der Reviews **überarbeiteten Ausarbeitung**

Zeitplan

- 24.10.2013 Vorbesprechung
- bis 07.11.2013 Themenauswahl und -zuordnung
- bis 21.11.2013 Abgabe der Gliederung
(letzte Möglichkeit zur Abmeldung vom Seminar in PAUL)
- bis 20.12.2013 Abgabe der schriftlichen Ausarbeitung
- bis 12.01.2014 Einreichen der Peer Reviews
- ab 27.01.2014 Blockseminar mit Vorträgen
- bis 16.02.2014 Abgabe der überarbeiteten Ausarbeitungen

Dokumente

- **Gliederung:**
 - Umfang 2-3 Seiten
 - Angabe des Titels und des Verfassers
 - Vorbemerkungen zu Thema und Konzeption
 - Kapitelüberschriften mit stichwortartiger Beschreibung des vorgesehenen Inhalts und des vorgesehenen Umfangs
 - Literaturverzeichnis mit Angabe der wichtigen Quellen
- **Ausarbeitung:**
 - Umfang 13-16 Seiten Text (ohne Titelseite, Bilder und Literaturverzeichnis)
 - Vorlagen für PdfLaTeX und Word werden auf der Seminar-Webseite bereitgestellt

Themenvorschläge (1)

- **Mining Sensor Data Streams**
Charu C. Aggarwal. Managing and Mining Sensor Data, 2013. Chapter 6.
- **Time Series and Multidimensional Streaming Outlier Detection**
Charu C. Aggarwal. Outlier Analysis, 2013. Chapter 8.
- **Outlier Detection in Discrete Sequences**
Charu C. Aggarwal. Outlier Analysis, 2013. Chapter 9.
- **One-class Classification**
Shehroz S. Khan and Michael G. Madden. A survey of recent trends in one class classification. AICS'09, 2009.
- **Introduction to Big Data Analytics**
Philip Russom. Big Data Analytics, TDWI Best Practices Report, 2011.

Themenvorschläge (2)

- **Log Analysis / Log Mining**

Adam Oliner, Archana Ganapathi, and Wei Xu. Advances and Challenges in Log Analysis. *Communications of the ACM*, 55(2), 2012.

- **OS Intrusion Detection**

S. Forrest, C. Warrender, and B. Pearlmutter. Detecting Intrusions using System Calls: Alternate Data Models, *IEEE ISRSP*, 1999.

- **ATM Fraud Detection**

WenTao Dong and YoungSung Soh. Image-based Fraud Detection in Automatic Teller Machine, *IJCSNS*, 6(11), 2006.

- **Credit Card Fraud Detection**

L. Delamaire, H. Abdou, and L. Pointon. Credit Card Fraud and Detection Techniques: a Review, *Banks and Bank Systems*, 4(2), pp.57-68, 2009.

- **Automatic model generation for monitoring process plants**

Oliver Niggemann, Benno Stein, Asmir Vodencarevic, Alexander Maier, and Hans Kleine Büning. Learning Behavior Models for Hybrid Timed Systems. *AAAI'12*, 2012.

Themenvorschläge (3)

- **PU Learning**

Bing Liu, Yang Dai, Xiaoli Li, Wee Sun Lee and Philip Yu. Building Text Classifiers Using Positive and Unlabeled Examples. ICDM'03, 2003.

- **One-class Support Vector Machines**

B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. Platt. Support Vector Method for Novelty Detection. Advances in Neural Information Processing Systems, 12(4), 2000.

- **RTI+**

Sicco Verwer, Mathijs de Weerdt, and Cees Witteveen. A Likelihood-ratio Test for Identifying Probabilistic Deterministic Real-time Automata from Positive Data. ICGI'10, 2010.

- **Feature Selection for One-class Classification**

Young-Seon Jeong, In-Ho Kang, Myong-Kee Jeong, and Dongjoon Kong. A New Feature Selection Method for One-Class Classification Problems, IEEE Transactions on Systems, Man, and Cybernetics, 42(6), 2012.

Wie geht's weiter?

- Einlesen in die relevante Literatur
 - Papers zu den Themenvorschlägen
 - Bücher von Charu Aggarwal: <http://charuaggarwal.net>
(online verfügbar via „Springer Link“)
 - Literatursuche: Google Scholar, ACM Digital Library, IEEE Xplore, DBLP, ...
- Ein Thema auswählen oder ein eigenes Thema vorschlagen
(eigene Themenvorschläge mit dem Dozenten besprechen)
- Bis 07.11.2013 verbindliche Zusage für ein Thema per Email
an Maik Anderka
 - first come first serve
 - Aktuelle Themenzuordnung auf der Seminar-Webseite
- Fragen, Probleme, Themenvorschläge
 - Maik Anderka kontaktieren