

# Vorbesprechung: Maschinelles Lernen und Experimentelles Design

Caroline Sporleder

Computational Linguistics  
Universität des Saarlandes

Sommersemester 2011

14.04.2011

# Organisatorisches

## Kein Weg drumrum:

Seit Mitte der 1990er Jahre ist die Computerlinguistik dominiert durch statistische und maschinelle Lernmethoden (ML).

State-of-the-Art Publikationen zu verstehen und zu bewerten setzt Kenntnisse von ML voraus.

## Seminarziele:

- 1 fachlich
  - Erlernen grundlegender Lernverfahren
  - Erwerben von Kenntnissen des experimentellen Designs im Bereich ML (Fragestellungen, Versuchsaufbau, Evaluation)
  - Kennenlernen von Anwendungsbeispielen in der CL  
⇒ am Ende des Seminars sollt ihr in der Lage sein die Grundzüge des ML zu verstehen und aktuelle Publikationen zu dem Thema kritisch evaluieren können
- 2 “transferrable skills”: Grundlagen des wissenschaftl. Arbeitens
  - Informationssuche & Aufbau des Gebiets (Konferenzen, Zeitschriften)
  - Lesen wissensch. Aufsätze
  - wissensch. Präsentationen halten
  - Hausarbeiten und andere wissensch. Arbeiten schreiben

Credits: 5 CPs (150 Std.)

gibt es für:

- Präsentation (10-15 Minuten)
- Hausarbeit (10-15 Seiten)
- Mitarbeit (setzt auch regelmäßige Teilnahme voraus!)

Gewichtung

- 50% Referat und Mitarbeit
- 50% Hausarbeit

- anspruchsvolles Thema  $\Rightarrow$  mit Arbeit verbunden
- etwas Statistik / Mathematik (sollte aber machbar sein)
- Feedback erwünscht

# Inhaltliches

## Hintergrundliteratur

- Tom Mitchell: *Machine Learning*, Boston: McGraw-Hill, 1997.
- Ian H. Witten & Eibe Frank: *Data Mining: Practical Machine Learning Tools and Techniques*, Amsterdam: Elsevier, 2005.
- Christopher Manning & Hinrich Schütze: *Foundations of Statistical Language Processing*, Massachusetts: MIT Press, 2003.

⇒ Semesterapparat (Kapitel werden noch angegeben)

## Weitere Literatur

- ausgewählte Aufsätze (s. Webseite)
- die allermeisten auf Englisch

# Semesterplan (vorläufig) (1)

- 14.4. Vorbesprechung, Organisatorisches
- 21.4. Einführung, einfache Verfahren
- 28.4. Einführung Information Theory  
Decision Tree Learning
- 5.5. Vertiefung Decision Tree Learning  
Lesen wissenschaftlicher Texte  
Halten eines wiss. Referats
- 12.5. Einführung in die Wahrscheinlichkeitstheorie  
Naive Bayes Classifier  
Referat: Decision Tree Learning
- 19.5. Evaluation von maschinellen Lernverfahren (1)  
Referat: Naive Bayes
- 26.5. Evaluation von maschinellen Lernverfahren (2)  
Referat: Evaluation
- 2.6. Himmelfahrt

- 9.6. Memory-Based Learning
- 16.6. Genetische Algorithmen  
Schreiben wissenschaftlicher Arbeiten  
Referat: Memory-based Learning
- 23.6. Fronleichnam
- 30.6. Feature Engineering und Example Selection  
Referat: Genetische Algorithmen
- 7.7. weitere Vorträge  
(z.B. Co-Training, Self-Learning, Ensemble Methods)
- 14.7. Zusammenfassung
- 21.7. fällt evtl. aus wg. Klausuren

- jeweils ein CL-Anwendungsbeispiel eines ML-Verfahrens
- Themen werden nächste Woche verteilt
- Auswahl der Aufsätze (normalerweise einer pro Referat) ist euch selbst überlassen
- Referate bitte mit mir vorbesprechen (setzt frühzeitige Planung voraus)