

Kern-Methoden zur Extraktion von Informationen

Sebastian Marius Kirsch
skirsch@moebius.inka.de



Back

Close

Gliederung

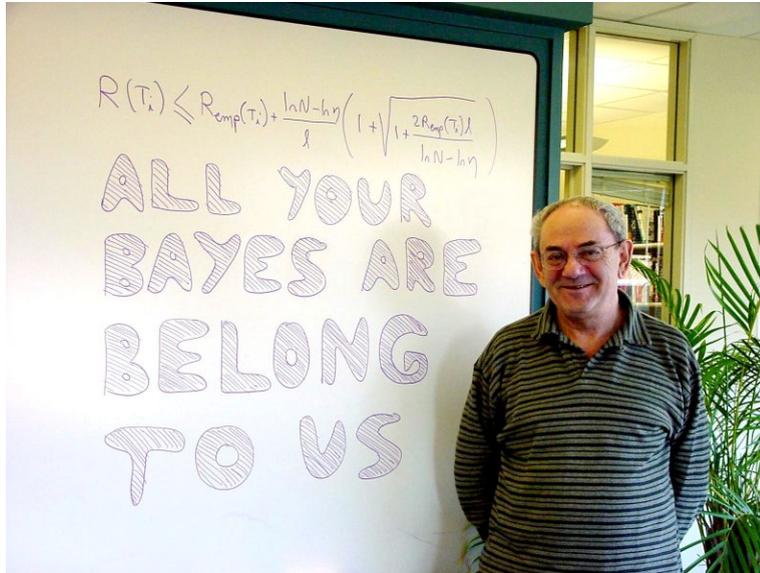
1. Verfahren zur Extraktion von Informationen
2. Extraktion von Beziehungen
3. Maschinelles Lernen auf Texten
4. Lernen mit Kernfunktionen
5. Kernfunktionen auf Bäumen
6. Experimente zur Extraktion von Beziehungen
7. Zusammenfassung



Back

Close

Vladimir Vapnik



„Vater“ der statistischen Lerntheorie und der *Support Vector Machine*, Prof. an der Royal Holloway University of London



Back

Close

Extraktion von Informationen

Ansätze:

- Ad-hoc-Verfahren (endliche Automaten, reguläre Ausdrücke)
von Hand erzeugte Muster, nach denen der Text abgesucht wird
- generative Modelle (Hidden Markov Models, Conditional Random Fields)
aus Beispielen Modell für die Erzeugung eines Textes generieren und daraus auf die Funktion seiner Bestandteile schließen
- diskriminative Methoden („maschinelles Lernen“, Support Vector Machines, Voted Perceptron)
Unterscheiden zwischen Beispielen mit und ohne der gewünschten Information



Back

Close

Extraktion von Beziehungen

- Stelle Beziehung zwischen Bestandteilen eines Satzes fest
- Ist eine Person, die im Satz genannt wird, bei einer ebenfalls genannten genannten Firma angestellt? (*member-affiliation*)
- Ist eine genannte Firma an einem ebenfalls genannten Ort ansässig? (*organization-location*)
- Mögliche Rollenfüller oft schon bekannt (*named entity extraction*)
- Unbekannt ist, ob sie etwas miteinander zu tun haben



Back

Close

Säcke voll Wörter: *bag-of-words*-Modell

- Standardmodell für Texte bei maschinellen Lernverfahren
- Worthäufigkeiten werden in einem Vektor gespeichert

„John Smith is the chief scientist of Hardcom Corp.“

↓

(..., 0, $\underbrace{1}_{\text{John Smith}}$, $\underbrace{1}_{\text{is}}$, 0, $\underbrace{1}_{\text{the}}$, $\underbrace{1}_{\text{chief}}$, $\underbrace{1}_{\text{scientist}}$, $\underbrace{1}_{\text{of}}$, $\underbrace{1}_{\text{Hardcom Corp.}}$, 0, ...)

„John Smith is meeting the chief scientist of Hardcom Corp.“

↓

(..., 0, $\underbrace{1}_{\text{John Smith}}$, $\underbrace{1}_{\text{is}}$, $\underbrace{1}_{\text{meeting}}$, $\underbrace{1}_{\text{the}}$, $\underbrace{1}_{\text{chief}}$, $\underbrace{1}_{\text{scientist}}$, $\underbrace{1}_{\text{of}}$, $\underbrace{1}_{\text{Hardcom Corp.}}$, 0, ...)

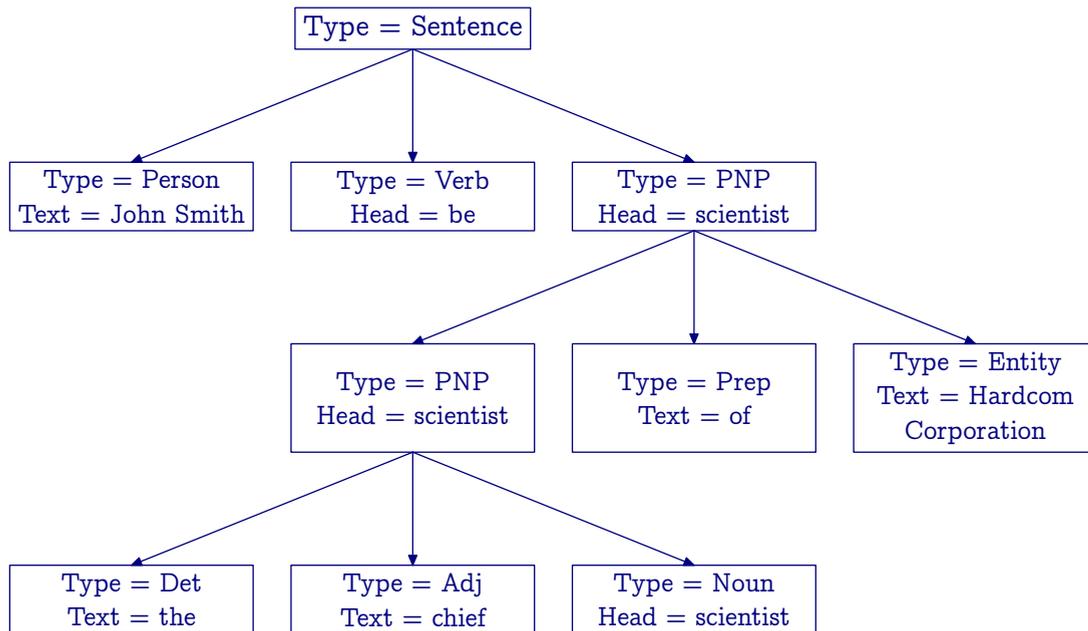


Back

Close

Flache Syntaxbäume

Flacher Syntaxbaum für den Satz „John Smith is the chief scientist of the Hardcom Corporation.“:



Back

Close

bag-of-words gegenüber Bäumen

bag-of-words:

- Leicht mit gängigen maschinellen Lernverfahren zu verarbeiten
- gute Erfolge bei der Kategorisierung von längeren Texten
- Struktur der Sätze geht verloren
- für einzelne Sätze und damit für Extraktion von Beziehungen ungeeignet

Flache Syntaxbäume:

- Standardrepräsentation in der Computerlinguistik
- Lassen Rückschlüsse auf Struktur des Satzes zu
- Mit maschinellen Lernverfahren üblicherweise nicht zu verarbeiten



Back

Close

Machinelles Lernen auf Bäumen?

Konventionelle Methode:

- Bestimmte Strukturen in Syntaxbaum suchen und in Vektor notieren, ob diese vorkommen
- Diesen Vektor als Eingabe für Lernverfahren benutzen

Andere Möglichkeit: Kern-Methoden (Müller u. a., 2001)

- arbeiten direkt auf Baum-Repräsentation
- zwei Komponenten:
 - Lernverfahren, das Kernfunktionen einsetzt (unabhängig von Eingabeformat und Aufgabe)
 - Kernfunktion für die Eingabe-Repräsentation (abhängig von Eingabeformat und Aufgabe)

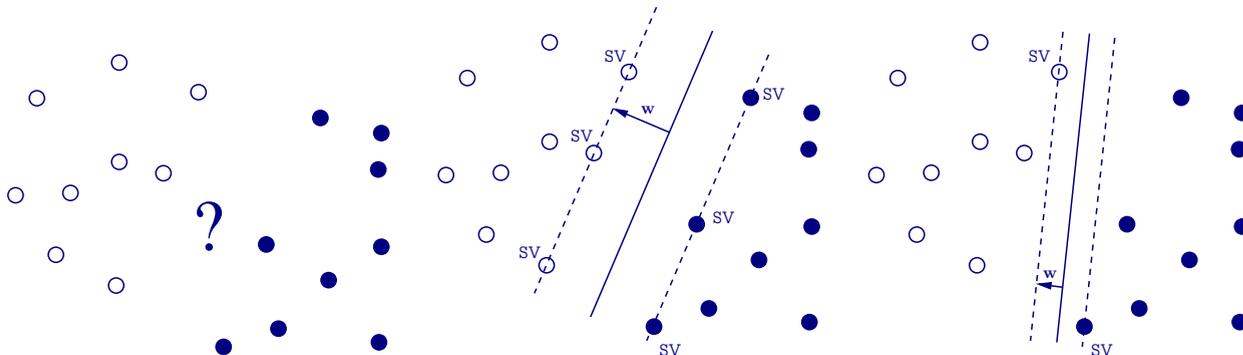


Back

Close

Lernverfahren mit Kernfunktionen

- *Support Vector Machine* versucht, Beispiele durch Hyper-ebene mit maximalem Abstand zu den beiden Klassen zu trennen
- Dadurch gute Regularisierung



(a) Aufgabe

(b) Lösung mit breitem Rand

(c) Lösung mit schmalen Rand



Kernfunktionen

- Beobachtung: Bestimmte Lernverfahren auf Vektoren lassen sich so formulieren, dass sie nur Skalarprodukte benutzen
- Auf Bäumen: Merkmale in Vektor extrahieren und dann Vektoren multiplizieren
- Gibt es eine Abkürzung?
- Kernfunktionen berechnen implizit das Skalarprodukt zwischen Vektoren mit extrahierten Merkmalen
- Jede symmetrische und positiv-definite Funktion ist eine Kernfunktion.
- Zahl der implizit verwendeten Merkmale kann sehr hoch (möglicherweise unendlich gross) sein.



Back

Close

Kernfunktionen auf Bäumen

- Kernfunktion dient als verallgemeinertes Abstandsmass
- soll Ähnlichkeit zwischen Bäumen ausdrücken
- Beispiel für Bäume: Zähle gemeinsam vorkommende Unterbäume; wird eingesetzt beim Parsing, um korrekt geparste Sätze von inkorrekten Sätzen zu unterscheiden. (Collins und Duffy, 2001)
- Für Extraktion von Beziehungen: Bestimme Unterbäume, die in syntaktischer Kategorie und möglicher Rolle übereinstimmen, und zähle bei diesen Unterbäumen, auf wie vielen Wörtern sie übereinstimmen. (Zelenko u. a., 2003)
- Dabei sind Einschübe in den Unterbäumen erlaubt (z. B. durch eingeschobene Nebensätze oder zusätzliche Attribute)
- Einschübe werden als geringere Ähnlichkeit gewertet



Back

Close

Experimentelle Ergebnisse

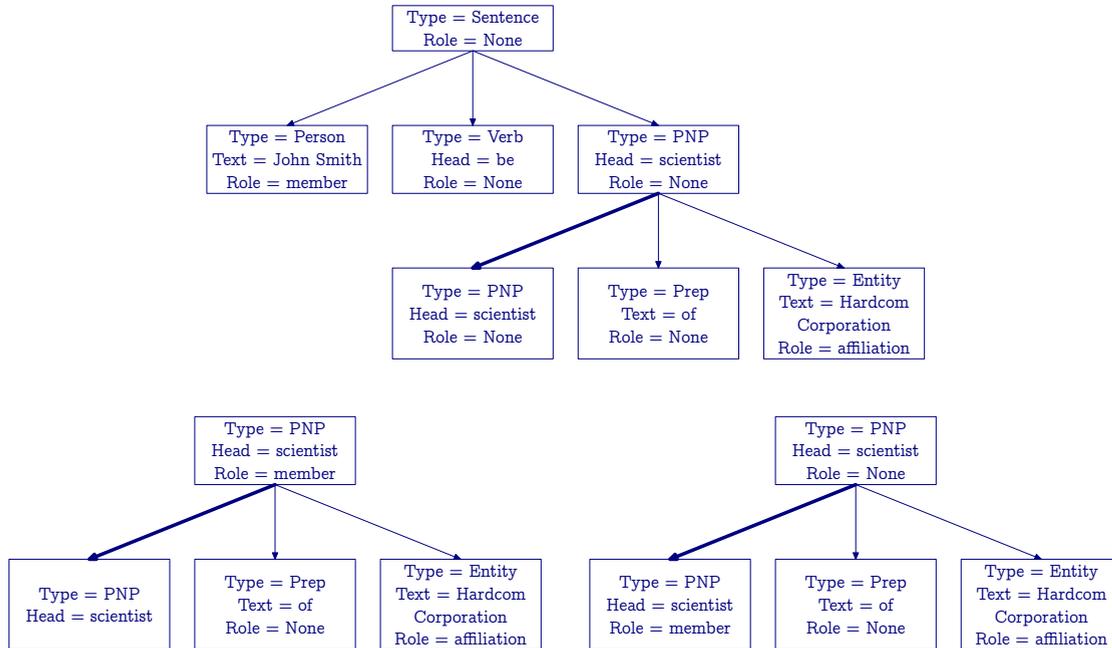
- Korpus von 200 Nachrichtenbeiträgen, selbst in flache Syntaxbäume überführt. (Zelenko u. a., 2003)
- Kandidaten für Beziehungen extrahiert und von Hand bewertet
- Zwei Lernverfahren auf Basis von Kernfunktionen: *Support Vector Machines* und *Voted Perceptron*
- Kernfunktion mit erlaubten Einschüben und ohne
- Als Vergleichsmaßstab: Drei herkömmliche Verfahren mit von Hand extrahierten Merkmalen (Winnow, *Naïve Bayes* und *Support Vector Machine* ohne Kernfunktion)



Back

Close

Kandidaten für Beziehungen



Performanz

Performanz für die Beziehung *person-affiliation*

	<i>recall</i> [%]	<i>precision</i> [%]
Naive Bayes	75,59	91,88
Winnnow	80,87	88,42
SVM (merkmalsbasiert)	76,21	91,67
<i>Voted Perceptron</i> (ohne Lücken)	79,58	89,74
SVM (ohne Lücken)	79,78	89,9
<i>Voted Perceptron</i> (mit Lücken)	81,62	90,05
SVM (mit Lücken)	82,73	91,32

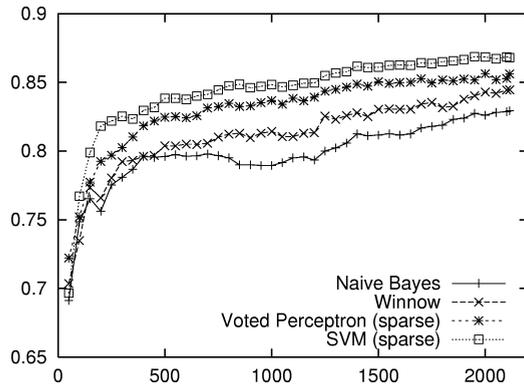


Back

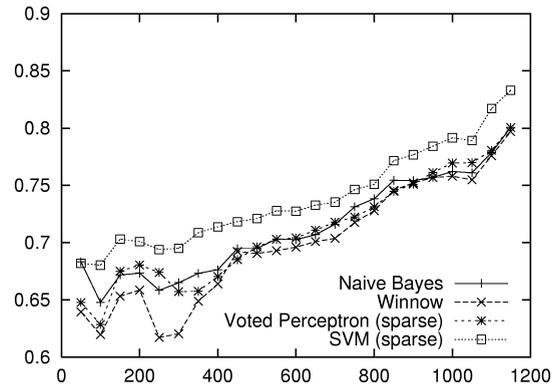
Close

Lernkurven

Merkmalsbasierte Methoden im Vergleich zu Kernmethoden



(g) *person-affiliation*



(h) *organization-location*



Back

Close

Zusammenfassung

- Diskriminative Methoden sind ein wichtiges Instrument zur Extraktion von Informationen.
- Mit Kernfunktionen können Standardrepräsentationen der Computerlinguistik direkt verarbeitet werden.
- Die Performanz auf dem gewählten Korpus übertrifft von Hand extrahierte Merkmale.
- Bei Übertragung auf eine andere Domäne sind nur geringe Anpassungen notwendig.



Back

Close

Ausarbeitung

Die Ausarbeitung zu diesem Vortrag ist verfügbar unter

<http://sites.inka.de/moebius/docs/semie-ho.pdf>

Vielen Dank für die Aufmerksamkeit!



Back

Close

Literatur

- [Collins und Duffy 2001] COLLINS, Michael ; DUFFY, Nigel: Convolution kernels for natural language. In: *Proceedings of NIPS-2001*, 2001
- [Müller u. a. 2001] MÜLLER, Klaus-Robert ; MIKA, Sebastian ; RÄTSCH, Gunnar ; TSUDA, Koji ; SCHÖLKOPF, Bernhard: An introduction to kernel-based learning algorithms. In: *IEEE Transactions on Neural Networks* 12 (2001), Nr. 2, S. 181–202
- [Zelenko u. a. 2003] ZELENKO, Dimitri ; AONE, Chinatsu ; RICARDELLA, Anthony: Kernel methods for relation extraction. In: *Journal of Machine Learning Research* (2003), Nr. 3, S. 1083–1106

