

Einführung in die Computerlinguistik

Statistische Verfahren in der lexikalischen Semantik

WS 2012/2013
Manfred Pinkal

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Akkuratheit

- Akkuratheit (engl. *accuracy*) ist das einfachste Maß:

Akkuratheit = korrekt klassifizierte Instanzen/alle Instanzen

- Fehlerrate (engl. *error rate*) ist der Komplementärbegriff zu Akkuratheit:

Fehlerrate = 1 – Akkuratheit

- Das Akkuratheitsmaß verdeckt oft tatsächlich relevante Eigenschaften eines Modells.

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Evaluation

- Annotation eines „Goldstandard“: Testkorpus mit der relevanten Zielinformation (z.B. Wortart)
- Automatische Annotation des Testkorpus mit statistischem Modell/ Klassifikator
- Messung der Performanz durch Vergleich von automatischer Annotation mit Goldstandard

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Konfusionsmatrix

- Fehlertypen für ADJA-Klassifikation:

| | | |
|-------------------------|------------------------|--------------------|
| | Echtes ADJA | Echtes NADJA |
| Klassifiziert als ADJA | ok | Korrektheitsfehler |
| Klassifiziert als NADJA | Vollständigkeitsfehler | ok |

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Konfusionsmatrix

- Fehlertypen für ADJA-Klassifikation:

| | Echtes ADJA | Echtes NADJA |
|-------------------------|----------------|----------------|
| Klassifiziert als ADJA | true positive | false positive |
| Klassifiziert als NADJA | false negative | true negative |

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Konfusionsmatrix

- (Fiktives) Beispiel:

| | Echtes ADJA | Echtes NADJA |
|-------------------------|-------------|--------------|
| Klassifiziert als ADJA | 20 | 80 |
| Klassifiziert als NADJA | 20 | 880 |

- Von insgesamt 1000 Fällen sind 900 korrekt (Wahre Positive und wahre Negative): Akkuratheit ist also 90%, Fehlerrate 10%.
- Tatsächlich ist die Adjektiverkennung miserabel: von fünf als ADJA klassifizierten Instanzen ist nur eine korrekt.
- Wir bestimmen **Recall** und **Präzision/ Precision** als klassenspezifische Maße, die Vollständigkeits- und Korrektheitsfehler (für eine gegebene Klasse) separat messen.

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Recall

| | Echtes ADJA | Echtes NADJA |
|-------------------------|----------------|----------------|
| Klassifiziert als ADJA | True positive | False positive |
| Klassifiziert als NADJA | False negative | True negative |

- Welcher Anteil der echten X wurde tatsächlich gefunden (als X klassifiziert)?

$$\text{Recall} = \text{True positives} / (\text{True positives} + \text{False negatives})$$

| | Echtes ADJA | Echtes NADJA |
|-------------------------|-------------|--------------|
| Klassifiziert als ADJA | 20 | 80 |
| Klassifiziert als NADJA | 20 | 880 |

$$\text{Recall für ADJA} = 20 / (20 + 20) = 0,5$$

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Präzision

| | Echtes ADJA | Echtes NADJA |
|-------------------------|----------------|----------------|
| Klassifiziert als ADJA | True positive | False positive |
| Klassifiziert als NADJA | False negative | True negative |

- Welcher Anteil der als X klassifizierten Instanzen ist tatsächlich ein X?

$$\text{Precision} = \text{True positives} / (\text{True positives} + \text{False positives})$$

| | Echtes ADJA | Echtes NADJA |
|-------------------------|-------------|--------------|
| Klassifiziert als ADJA | 20 | 80 |
| Klassifiziert als NADJA | 20 | 880 |

$$\text{Precision für ADJA} = 20 / (20 + 80) = 0,2$$

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Präzision und Recall

- Präzision und Recall sind im allgemeinen nur zusammen aussagekräftig
 - Hohe Präzision, hoher Recall: gutes Modell
 - Niedrige Präzision, niedriger Recall: schlechtes Modell
 - Hohe Präzision, niedriger Recall: „Vorsichtiges“ Modell
 - Findet nicht alle Instanzen von X
 - Klassifiziert kaum keine Nicht-Xe als X
 - Niedrige Präzision, hoher Recall: „Mutiges“ Modell
 - Findet fast alle Instanzen von X
 - Klassifiziert viele nicht-Xe fehlerhaft als X
 - Extremfälle
 - Modell klassifiziert alles als X: Recall 100%, Precision niedrig
 - Modell klassifiziert nichts als X: Recall 0%, Precision nicht definiert

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UdS Computerlinguistik

Wortart-Tagging

- Standard Wortart-Tagger arbeiten mit ca. 50 Klassen und haben dabei eine Akkuratheit von deutlich über 99%.
- Sie gehen dabei etwas anders vor, als hier demonstriert: Sie verwenden maschinelle Lernverfahren, die nicht nur die besten POS-Tags für die einzelnen Wörter im Satz, sondern die beste POS-Kette für einen ganzen Satz zu bestimmen versuchen.
- Beispiel: Auch wenn in „*I made her duck*“ die wahrscheinlichste Wortart für *her* Personalpronomen und für *duck* Gattungssubstantiv ist, ist die Kombination der Wortarten ziemlich unwahrscheinlich.
- Die Methode, beste Wahrscheinlichkeiten für Sequenzen zu von Wortart-Tags zu bestimmen, sind „Hidden Markov Models“ (HMMs) – wir kommen später darauf zurück.

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UdS Computerlinguistik

F-Score

- Der „F-Score“ ist ein Maß für die „Gesamtgüte“ der Klassifikation, in das Precision und Recall eingehen.

$$F = \frac{2PR}{P + R}$$

- F-Score für die Klasse ADJA im Beispiel:

$$F = \frac{2 * 0,2 * 0,5}{0,2 + 0,5} = 0,29$$

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UdS Computerlinguistik

Lexikalische Mehrdeutigkeit

- Äußerungs- und Textverstehen impliziert die Erkennung der korrekten, im Kontext intendierten Äußerungsbedeutung.
- Wörter sind vielfach mehrdeutig:
 - *Bank*: *Geldinstitut* / *Sitzmöbel*
 - *Maschine*: *Flugzeug* / *Motorrad* / *Technisches Gerät*
 - *Absatz*: *Schuh* / *Treppe* / *Text* / *Verkauf*
 - *aufgeben*: *einen Plan* / *einen Koffer aufgeben*
- Die Disambiguierung der Wortbedeutung (engl. "[Word-sense disambiguation](#)": WSD) ist zentrale Aufgabe der Computerlinguistik.

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UdS Computerlinguistik

WSD

Ich gehe heute einkaufen, und ich muss vorher zur Bank.

- Wissensbasierte Disambiguierung des Zielwortes *Bank* durch Inferenz mit Weltwissen:
 - Praktisch nicht machbar: Riesige Mengen an handkodiertem Weltwissen wären nötig.
- Alternative: Statistische Modellierung
 - Identifikation von wortspezifischen Merkmalen ("in Objektposition des Verbs *müssen*", "*einkaufen* als Kontextwort")
 - Nicht sinnvoll, da es tausende von mehrdeutigen Wörtern gibt, die keinem gemeinsamen Muster folgen.

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UdS Computerlinguistik

Statistische Modellierung: Allgemeines Schema

- Manuelle Korpusannotation
- Merkmalspezifikation
- Automatische Merkmalsextraktion
- Bestimmung und Training eines statistischen Modells
- Evaluierung

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UdS Computerlinguistik

WSD: Korpusannotation

- Korpusannotation:
 - Spezifikation des "Annotationsschemas": Übernahme von Wortbedeutungen aus einem Wörterbuch oder Thesaurus (Standard: WordNet-Synsets)
 - Annotation aller Zielwort-Instanzen mit einer Wortbedeutung

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UdS Computerlinguistik

Trainings-Korpus

- ...
- (A237) ... Für diejenigen, denen Komfort wichtig ist, haben wir eine Bank mit leicht schwingender Rückenlehne entwickelt. ...
- (A295) ... Ich suche noch eine Bank für meinen Garten und sondiere deshalb gerade Angebote. ...
- (A303) ... Habe im März 2000 einen höheren Betrag bei einer Bank angelegt. ...
- (A452) ... Beim Test Anlageberatung der Banken löste kein Institut die einfache Frage nach einer sicheren Anlage wirklich gut. ...
- ...

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UdS Computerlinguistik

Trainings-Korpus: Annotation mit WSD-Information

...

(A237) ... Für diejenigen, denen *Komfort wichtig* ist, haben wir eine Bank <bank1> mit leicht schwingender Rückenlehne entwickelt. ...

(A295) ... Ich *suche* noch eine Bank <bank1> für meinen *Garten* und *sondiere* deshalb gerade *Angebote*. ...

(A303) ... Habe im März 2000 einen höheren *Betrag* bei einer Bank<bank2> angelegt. ...

(A452) ... Beim *Test* *Anlageberatung* der Banken <bank2> löste kein Institut die einfache *Frage* nach einer *sicheren Anlage* wirklich gut. ...

...

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UdS Computerlinguistik

Zielwörter und Kontextwörter

...

(A237) ... Für diejenigen, denen *Komfort wichtig* ist, haben wir eine Bank <bank1> mit leicht schwingender Rückenlehne entwickelt. ...

(A295) ... Ich *suche* noch eine Bank <bank1> für meinen *Garten* und *sondiere* deshalb gerade *Angebote*. ...

(A303) ... Habe im März 2000 einen höheren *Betrag* bei einer Bank<bank2> angelegt. ...

(A452) ... Beim *Test* *Anlageberatung* der Banken <bank2> löste kein Institut die einfache *Frage* nach einer *sicheren Anlage* wirklich gut. ...

...

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UdS Computerlinguistik

WSD: Merkmalspezifikation

- Wir verwenden Kontextwörter als Merkmale für die Disambiguierung. Präziser ausgedrückt:
- Wir nehmen für alle Zielwörter eine gemeinsame Merkmalsmenge an: Vorkommen/ Nichtvorkommen der n (z.B. n=1000) häufigsten **Inhaltswörter** (Substantive, Verben, Adjektive) im Kontext des Zielwortes.
- Den Kontext einer Instanz legen wir als den Satz fest, in dem die Instanz vorkommt (alternativ: das Fenster mit fester Länge von k Wörtern rechts und links von der Instanz (z.B. k=5)).

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UdS Computerlinguistik

WSD: Merkmalsextraktion

- Wir **lemmatisieren** die Kontextwörter (Bestimmung des Wortstamms durch morphologische Analyse)
- Wir ermitteln für jedes Vorkommen eines Zielwortes ein spezifisches Merkmalsmuster v , indem wir für jedes i : $1 \leq i \leq 1000$ setzen:
 - $v_i = 1$, wenn das Wort w_i als Kontextwort im Satz auftritt
 - $v_i = 0$ sonst.
- Alle Merkmale sind Boole'sche Merkmale ($\in \{0,1\}$). Das Merkmalsmuster v kann als Vektor in einem hochdimensionalen Raum betrachtet werden.

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UdS Computerlinguistik

Merkmalsextraktion, Beispiel

| Instanz Nr | A237 | A295 | A303 | A452 | ... |
|------------|-------|-------|-------|-------|-----|
| Annotation | bank1 | bank1 | bank2 | bank2 | ... |
| Frage | 0 | 0 | 0 | 1 | ... |
| Komfort | 1 | 0 | 0 | 0 | ... |
| anlegen | 0 | 0 | 1 | 0 | ... |
| Betrag | 0 | 0 | 1 | 0 | ... |
| Garten | 0 | 1 | 0 | 0 | ... |
| suchen | 0 | 1 | 0 | 0 | ... |
| fahren | 0 | 0 | 0 | 0 | ... |
| richtig | 0 | 0 | 0 | 0 | ... |
| Test | 0 | 0 | 0 | 1 | ... |
| ... | ... | ... | ... | ... | ... |

Neue Instanz:

*Keine Frage: In
einen ordentlichen
Garten gehören
neben einer Bank
auch die richtigen
Möbel.*

| Neu |
|-----|
| ? |
| 1 |
| 0 |
| 0 |
| 0 |
| 1 |
| 0 |
| 0 |
| 1 |
| 0 |
| ... |

WSD: Statistisches Modell

- Wie bestimmen wir den Wortsinn einer neuen Instanz von *Bank* auf der Grundlage des Musters von Kontextwörtern v ?

Versuch: Analog zum POS-Modell der letzten Woche:

- Wir zählen zunächst für jedes Merkmalsmuster aus, wie oft im Trainingskorpus das Muster mit *bank1* und *bank2* vorkommt.
- Wir schätzen die bedingten Wahrscheinlichkeiten $P(\text{bank1}|v)$ und $P(\text{bank2}|v)$ auf der Grundlage dieser Frequenzen.

$$P(s|v) = \frac{P(s,v)}{P(v)} \approx \frac{Fr(s,v)}{Fr(v)}$$

- Wir weisen den wahrscheinlicheren Wortsinn zu.
- Sparse-Data-Problem:**
 - 1000 Wörter, je 2 Werte: 2^{1000} alternative Kontextmuster.

Das Bayessche Theorem

- Das Bayessche Theorem oder die Bayes-Regel:

$$P(E|F) = \frac{P(F|E) \cdot P(E)}{P(F)}$$

- Die Bayes-Regel ist ein elementares Gesetz der Wahrscheinlichkeitstheorie. Sie ist überall da nützlich, wo der Schluss von einer Größe F auf eine andere Größe E bestimmt werden soll (typischerweise von einem Symptom auf eine relevante Eigenschaft/ die Ursache), die Abhängigkeit in der anderen Richtung (von der Ursache auf das Symptom) aber besser zugänglich ist.

Bayes-Theorem und WSD

- Merkmalsmuster v : Symptom
- Wortsinn s : Ursache
- Mit Bayes-Regel : $P(s|v) = \frac{P(v|s) \cdot P(s)}{P(v)}$
- Der wahrscheinlichste Wortsinn: $\max_s P(s|v) = \max_s \frac{P(v|s) \cdot P(s)}{P(v)} = \max_s P(v|s) \cdot P(s)$
- $P(s)$ ist die globale, "a priori"-Wahrscheinlichkeit des Wortsinns s .
- $P(v)$, die Wahrscheinlichkeit des Merkmalsmusters, wird nicht mehr benötigt.
- Wie ermitteln wir $P(v|s)$? – Wiederum: **Sparse-Data-Problem:**
 - Die Auftretenshäufigkeit eines bestimmten Musters mit einer Lesart ist typischerweise klein (meistens sogar 0) und erlaubt deshalb keine verlässliche Abschätzung von $P(v|s)$.

Unabhängigkeitsannahme

- Unter der Voraussetzung, dass zwei Ereignisse E_1 und E_2 unabhängig sind, ist die gemeinsame Wahrscheinlichkeit von E_1 und E_2 das Produkt der Einzelwahrscheinlichkeiten:

$$P(E_1, E_2) = P(E_1) * P(E_2)$$

$$P(E_1, E_2 | F) = P(E_1 | F) * P(E_2 | F)$$

- Die Regel gilt für Produkte mit beliebig vielen Faktoren. Unter der ("naiven") Annahme, dass die Merkmale unabhängig voneinander auftreten, lässt sich die Wahrscheinlichkeit eines Merkmalsmusters approximieren als Produkt der Einzelwahrscheinlichkeiten für seine Komponenten:

$$P(v | s) \approx \prod_{v_i} P(v_i | s)$$

- Maschinelle Lernverfahren, die diese Unabhängigkeitsannahme nutzen, um Wahrscheinlichkeiten trotz geringer Datenmengen zu approximieren, heißen "Naive Bayes Classifier".

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

... zu Wahrscheinlichkeitsschätzungen von Kontextmerkmalen

- Wir gehen von insgesamt 500 Instanzen von "Bank" im Trainingskorpus aus, davon 200 als bank1 und 300 als bank2 annotiert.

| | bank1 | bank2 |
|---------|-------|-------|
| Frage | 11 | 15 |
| Komfort | 7 | 3 |
| anlegen | 3 | 84 |
| Betrag | 5 | 41 |
| Garten | 40 | 1 |
| suchen | 7 | 32 |
| fahren | 4 | 24 |
| richtig | 12 | 21 |
| Test | 2 | 5 |
| ... | ... | ... |

| | $P(I bank1)$ | $P(0 bank1)$ |
|---------|--------------|--------------|
| Frage | 0,055 | 0,945 |
| Komfort | 0,035 | 0,965 |
| anlegen | 0,015 | 0,985 |
| Betrag | 0,025 | 0,975 |
| Garten | 0,200 | 0,800 |
| suchen | 0,035 | 0,965 |
| fahren | 0,020 | 0,980 |
| richtig | 0,060 | 0,940 |
| Test | 0,010 | 0,990 |
| ... | ... | ... |

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Von binären Merkmalsmustern ...

| Instanz Nr | ... | A237 | A295 | A303 | A452 | ... |
|------------|-----|-------|-------|-------|-------|-----|
| Annotation | ... | bank1 | bank1 | bank2 | bank2 | ... |
| Frage | ... | 0 | 0 | 0 | 1 | ... |
| Komfort | ... | 1 | 0 | 0 | 0 | ... |
| anlegen | ... | 0 | 0 | 1 | 0 | ... |
| Betrag | ... | 0 | 0 | 1 | 0 | ... |
| Garten | ... | 0 | 1 | 0 | 0 | ... |
| suchen | ... | 0 | 1 | 0 | 0 | ... |
| fahren | ... | 0 | 0 | 0 | 0 | ... |
| richtig | ... | 0 | 0 | 0 | 0 | ... |
| Test | ... | 0 | 0 | 0 | 1 | ... |
| ... | ... | ... | ... | ... | ... | ... |

| | bank1 | bank2 |
|---------|-------|-------|
| Frage | 11 | 15 |
| Komfort | 7 | 3 |
| anlegen | 3 | 84 |
| Betrag | 5 | 41 |
| Garten | 40 | 1 |
| suchen | 7 | 32 |
| fahren | 4 | 24 |
| richtig | 12 | 21 |
| Test | 2 | 5 |
| ... | ... | ... |

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

... zu Wahrscheinlichkeitsschätzungen von Kontextmerkmalen

- Wir gehen von insgesamt 500 Instanzen von "Bank" im Trainingskorpus aus, davon 200 als bank1 und 300 als bank2 annotiert.

| | bank1 | bank2 |
|---------|-------|-------|
| Frage | 11 | 15 |
| Komfort | 7 | 3 |
| anlegen | 3 | 84 |
| Betrag | 5 | 41 |
| Garten | 40 | 1 |
| suchen | 7 | 32 |
| fahren | 4 | 24 |
| richtig | 12 | 21 |
| Test | 2 | 5 |
| ... | ... | ... |

| | $P(I bank2)$ | $P(0 bank2)$ |
|---------|--------------|--------------|
| Frage | 0,050 | 0,950 |
| Komfort | 0,010 | 0,990 |
| anlegen | 0,280 | 0,720 |
| Betrag | 0,137 | 0,863 |
| Garten | 0,003 | 0,997 |
| suchen | 0,107 | 0,893 |
| fahren | 0,080 | 0,920 |
| richtig | 0,070 | 0,930 |
| Test | 0,017 | 0,983 |
| ... | ... | ... |

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Beispiel

Keine Frage: In einen ordentlichen Garten gehören neben einer Bank auch die richtigen Möbel.

$$\max_s P(s | v) = \max_s P(v | s) \cdot P(s)$$

$$s \in \{bank1, bank2\}$$

$$P(v | bank1) \approx \prod_{v_i} P(v_i | bank1)$$

$$P(v | Bank2) \approx \prod_{v_i} P(v_i | Bank2)$$

| | $P(1 bank1)$ | $P(0 bank1)$ | v_i | $P(v_i bank1)$ |
|---------|--------------|--------------|-------|----------------|
| Frage | 0,055 | 0,945 | 1 | 0,055 |
| Komfort | 0,035 | 0,965 | 0 | 0,965 |
| anlegen | 0,015 | 0,985 | 0 | 0,985 |
| Betrag | 0,025 | 0,975 | 0 | 0,975 |
| Garten | 0,200 | 0,800 | 1 | 0,200 |
| suchen | 0,035 | 0,965 | 0 | 0,965 |
| fahren | 0,020 | 0,980 | 0 | 0,980 |
| richtig | 0,060 | 0,940 | 1 | 0,060 |
| Test | 0,010 | 0,990 | 0 | 0,990 |
| ... | ... | ... | ... | ... |

$$P(v | bank1) \approx \prod_{v_i} P(v_i | bank1) = 0,000572$$

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Beispiel

Keine Frage: In einen ordentlichen Garten gehören neben einer Bank auch die richtigen Möbel.

| | $P(1 bank2)$ | $P(0 bank2)$ | v_i | $P(v_i bank2)$ |
|---------|--------------|--------------|-------|----------------|
| Frage | 0,050 | 0,950 | 1 | 0,050 |
| Komfort | 0,010 | 0,990 | 0 | 0,990 |
| anlegen | 0,280 | 0,720 | 0 | 0,720 |
| Betrag | 0,137 | 0,863 | 0 | 0,863 |
| Garten | 0,003 | 0,997 | 1 | 0,003 |
| suchen | 0,107 | 0,893 | 0 | 0,893 |
| fahren | 0,080 | 0,920 | 0 | 0,920 |
| richtig | 0,070 | 0,930 | 1 | 0,070 |
| Test | 0,017 | 0,983 | 0 | 0,983 |
| ... | ... | ... | ... | ... |

$$P(v | Bank2) \approx \prod_{v_i} P(v_i | Bank2) = 0,000006$$

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

$$\max_s P(s | v) = \max_s P(v | s) \cdot P(s)$$

$$s \in \{bank1, bank2\}$$

$$P(v | bank1) \approx \prod_{v_i} P(v_i | bank1) = 0,000572 \quad P(bank1) = 0,4$$

$$P(v | bank2) \approx \prod_{v_i} P(v_i | bank2) = 0,000006 \quad P(bank2) = 0,6$$

$$P(v | bank1) \cdot P(bank1) = 0,000228$$

$$P(v | bank2) \cdot P(bank2) = 0,000004$$

$$\max_s P(s | v) = bank1$$

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

WSD

- Eine der schwierigsten Aufgaben in der Computerlinguistik:
- Sehr viele Wörter sind auf sehr unterschiedliche Weise mehrdeutig. Man benötigt riesige Mengen von Trainingsmaterial.
- Alle bisher vorgestellten Lernverfahren sind „überwachte“ (supervised) Lernverfahren: Sie erfordern die manuelle Annotation eines Trainingskorpus.
- Attraktiver sind „halbüberwachte“ (semi-supervised) Verfahren, bei denen ein großes Trainingskorpus (teil-)automatisch auf der Grundlage einer kleinen Menge von handannotierten „Seed-Daten“ erzeugt wird.
- Noch attraktiver sind „unüberwachte“ statistische Verfahren, die Resultate ohne jedes Training erzielen.