



Diskriminatives syntaktisches Reranking für **SMT**

Fortgeschrittene Themen der
statistischen maschinellen
Übersetzung

Janina Nikolic

Agenda

- Problem: Ranking des SMT Systems
- Lösung: Reranking-Modell
 - Nutzung Perzeptron-Algorithmus
 - Besonderheit: syntaktische Merkmale
- Evaluation

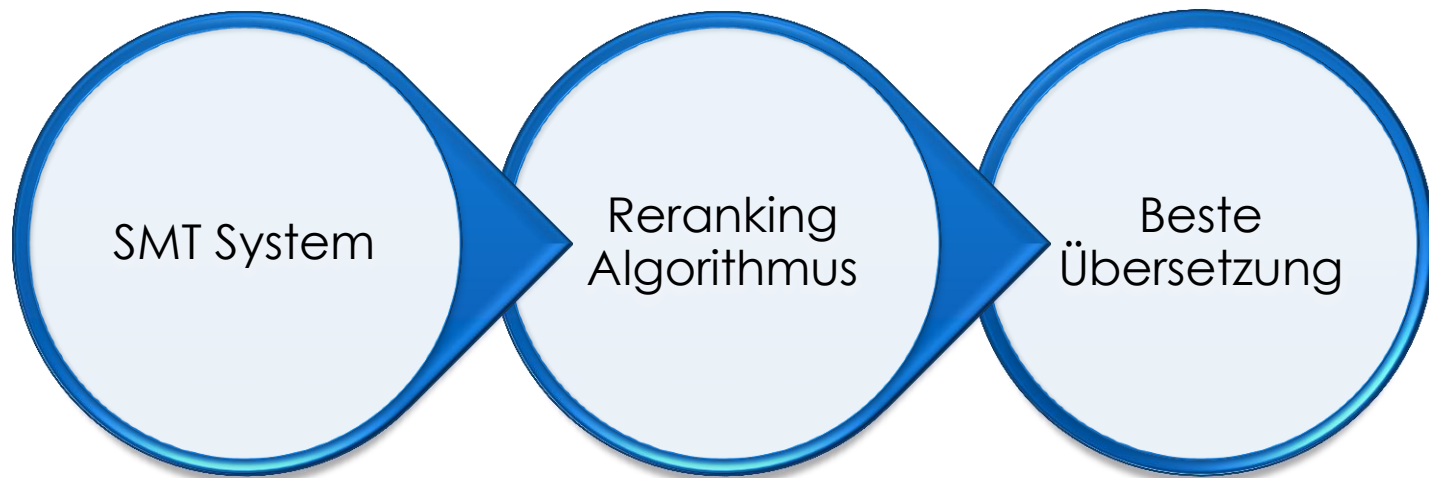
Problemstellung

- **Komponenten der SMT Systeme:**
Übersetzungsmodell, Sprachmodell,
Reordering Modell
- **Ziel:** Verbesserung des Sprachmodells
- **Ausgangspunkt:** SMT System liefert als Output
eine Liste der besten n Übersetzungen
- **Problem:** Ranking nicht immer das Beste

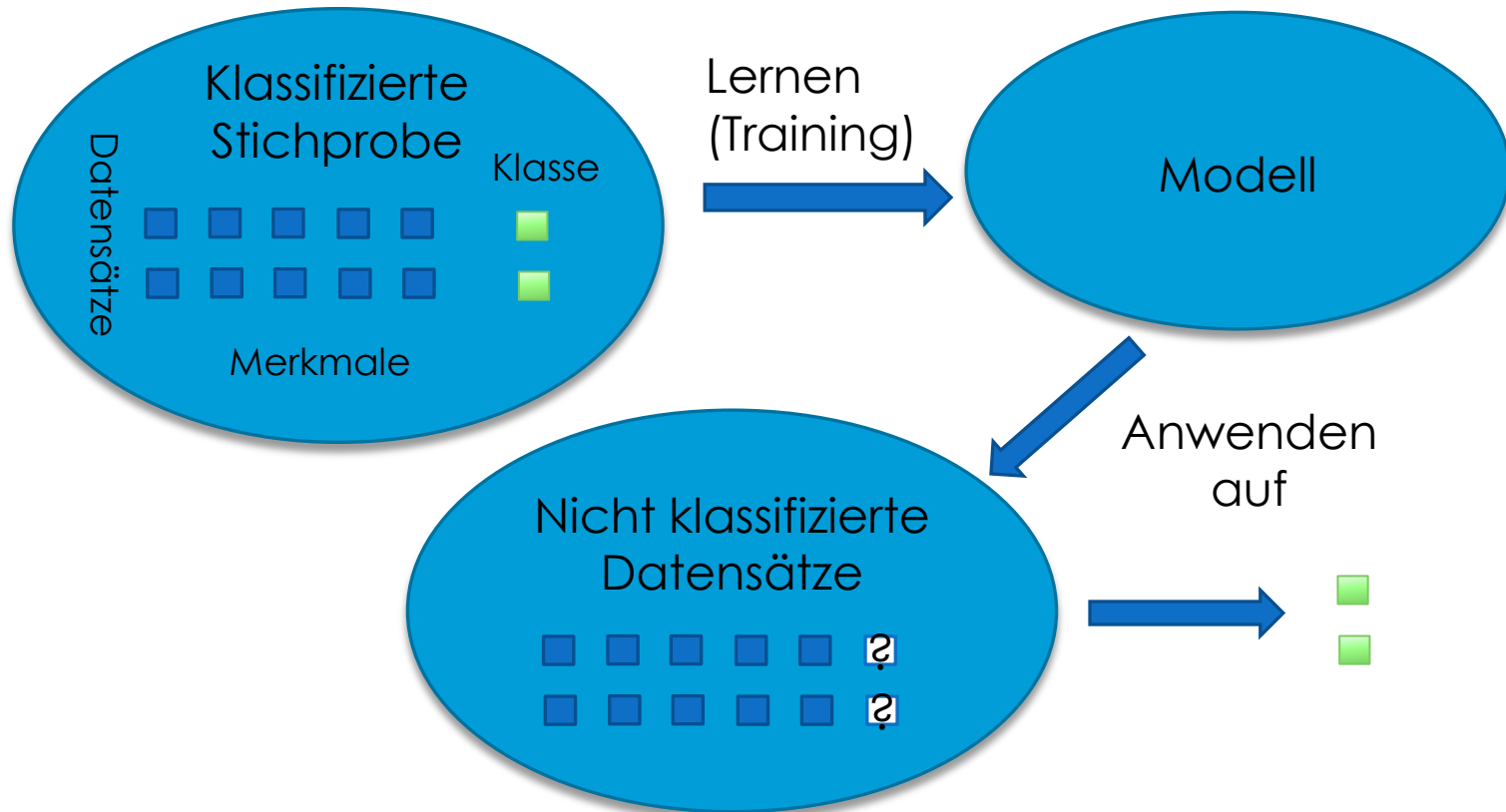
„Der Himmel wird morgen blau sein“

- The sky blue tomorrow
- The sky will be blue tomorrow
- Sky will be blue tomorrow

Lösung: Reranking



Reranking durch Klassifikation



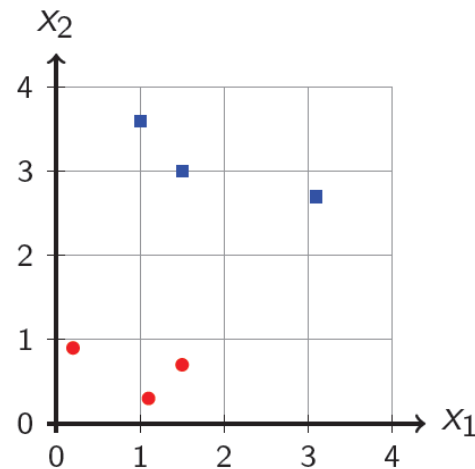
Was wird klassifiziert?

- Daten werden Klasse zugeordnet
- Daten: übersetzte Sätze aus der Liste (bzw. deren Merkmale)
- Entspricht Referenzübersetzung (oracle best)
→ Klasse 1, sonst Klasse 0
- Merkmale durch Merkmalsvektor repräsentiert
- Anpassbare Gewichtungen für jedes Merkmal (Gewichtungsvektor): Wie entscheidend ist ein Merkmal für eine gute Übersetzung?

Beispiel

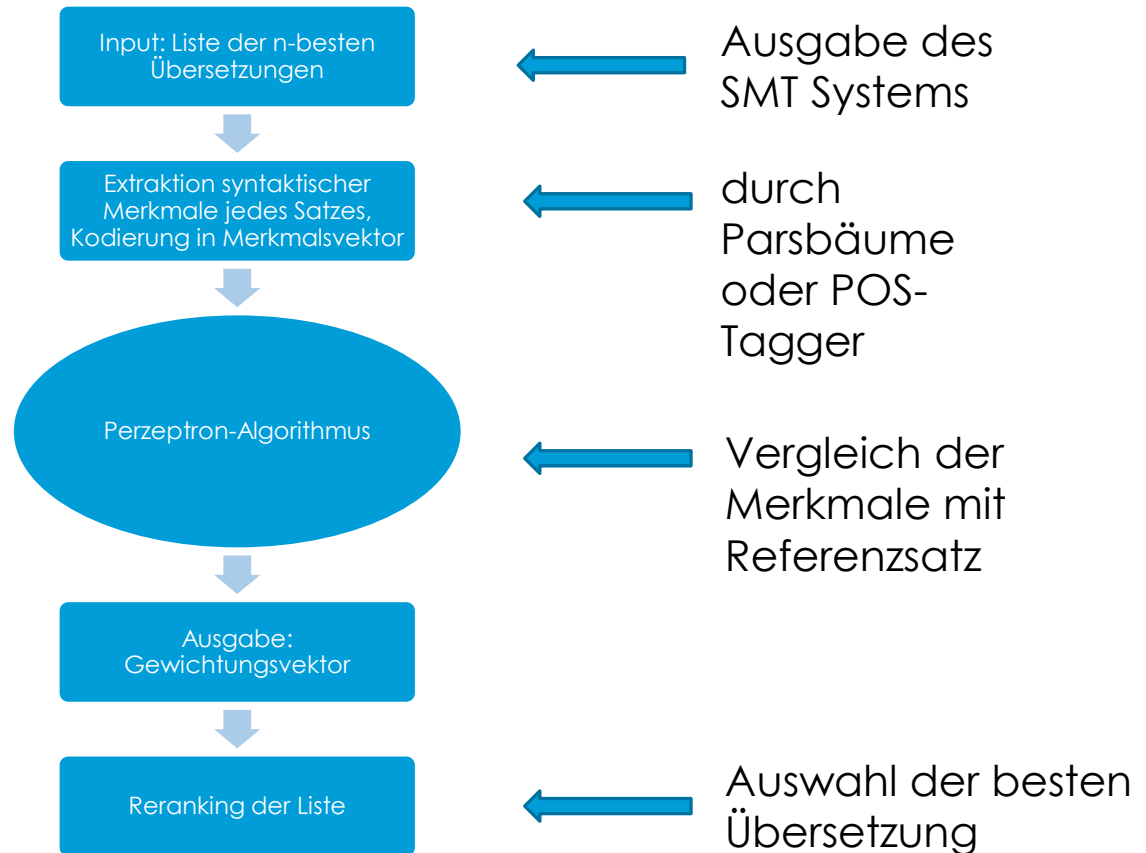
- Gegeben: Trainingsmenge

x_1	x_2	$c(\mathbf{x})$
1.5	0.7	●
0.2	0.9	●
1.1	0.3	●
1	3.6	■
3.1	2.7	■
1.5	3	■



- Gesucht: Gewichtsvektor, der die Trainingsmenge korrekt klassifiziert

Diskriminatives Sprachmodell (DLM)



Merkmalsvektoren

- Beschreiben die syntaktischen Eigenschaften der Sätze: enthält ein Verb, Verb Agreement etc.
- Trainingsdaten, mit denen Modell gelernt wird
- Testdaten, die klassifiziert werden
- Jeder Merkmalsvektor gehört ausschließlich zu einer Klasse

Wie werden die
syntaktischen Merkmale
extrahiert?

Merkmalsextraktion

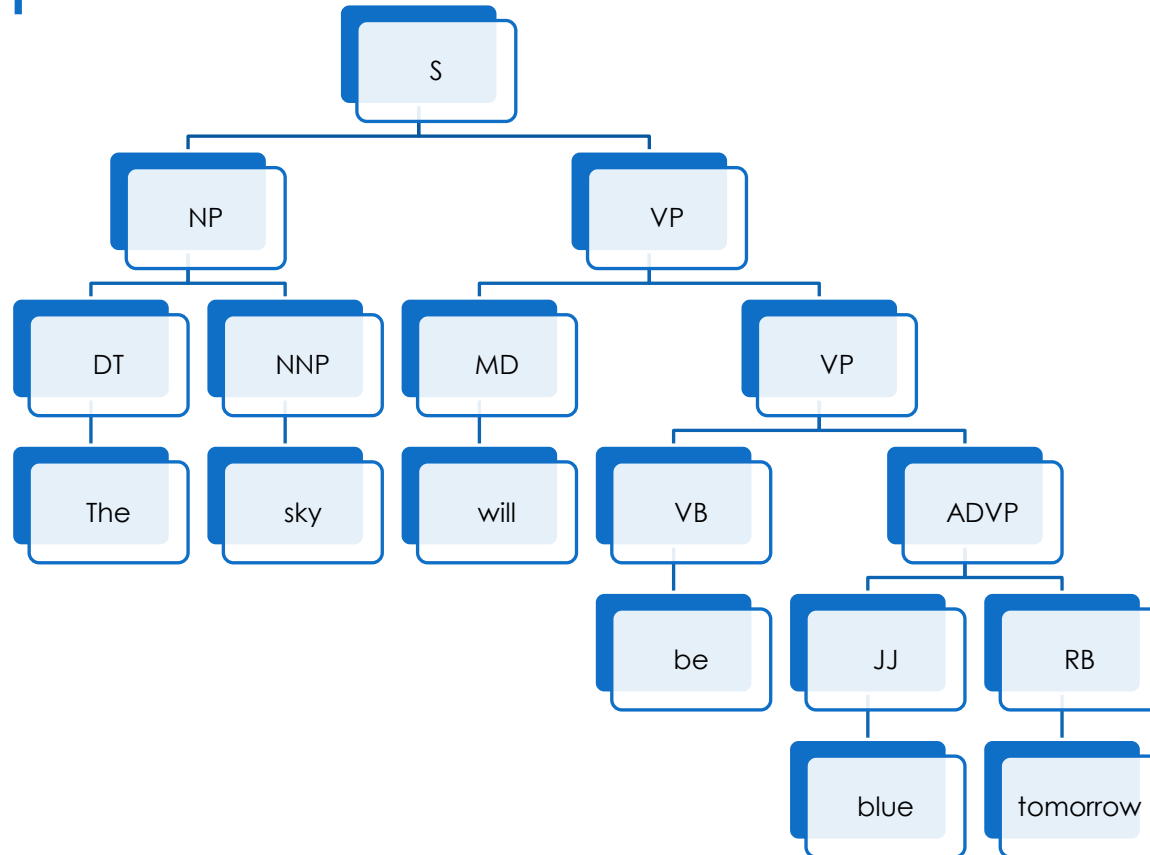
Vollständige
Parabäume

POS
Sequenzen

Parsbäume

- Drei Arten von syntaktischen Informationen werden extrahiert
 - Informationen über Sequenzen
 - Head-Informationen (Head = Kern einer Phrase)
 - Kontextfreie Grammatikregeln (Informationen über NT-Folgen)

Beispiel Parsbaum



Unterscheidung von Sequenzen

- POS

The/DT sky/NNP will/MD be/VB blue/JJ
tomorrow/RB

- SEQ-B: Teilstrukturen (chunks)

The/NP^b sky/NP^c will/VP^b be/VP^b blue/ADV^b
tomorrow/ADV^c

- SEQ-C: Kombination der ersten beiden Strukturen

The/DT-NP^b ...

POS Tagger

- Zwei Tagging-Ansätze
 - Conditional Random Fields
 - Einfacher Unigramm-Tagger
 - Nutzt keinen Kontext
 - <UNK> als Tag für unbekannte Wörter

Beispiele für Merkmale

- Abfolgen von POS-Tags
- Häufigkeit von POS-Typen für eine bestimmte Satzlänge:
 $\text{length}(x)/\text{num}(\text{POS},x)$
- Fehlen gewisser POS-Typen (z.B. Verb)
- Verb Agreement („George was shouting and screaming“)

Merkmalsextraktion
erfolgreich – Was nun?

Perzeptron-Algorithmus

- Künstliches neurales Netz mit anpassbaren Gewichtungen und Schwellwert
- Bildet biologisches Nervensystem nach (ein Neuron)
- Eingabe: Merkmalsvektor eines übersetzten Satzes
- Aufgabe: Lernen des optimalen Gewichtungsvektors aus den Trainingsbeispielen

Perzeptron-Algorithmus

- Zu Beginn Gewichtungsvektor 0
- Bestimmung der besten Übersetzung (oracle best) hinsichtlich BLEU
- ausgewählter Satz aus der Liste: Skalarprodukt aus Merkmalsvektor und Gewichtungsvektor
- Wenn ausgewählter Satz mit oracle best Satz übereinstimmt, fertig (Wert größer als Schwellenwert)
- Sonst: Gewichtungsvektor wird angepasst

Perzeptron-Algorithmus

y^i : oracle best

z^i : ausgewählter Satz der Liste

$$z^i = \phi(z) * \omega$$

Merkmalsvektor

Gewichtsvektor

wenn $z^i \neq y^i$  Anpassung Gewichtsvektor

Anpassung Gewichtsvektor

$$\omega = \omega + \underbrace{\phi(y^i) - \phi(z^i)}_{\text{Änderung: gewünschte Ausgabe - tatsächliche Ausgabe}}$$

Neue Gewichtung

Alte Gewichtung

Reranking

- Jeder Satz der n-best Liste erhält einen neuen Score, der sich folgendermaßen ergibt:
- Gewichteter Score des SMT-Systems für den Satz + Gewichtungsvektor * Merkmalsvektor
- Satz mit dem höchsten Score wird als beste Übersetzung ausgewählt

Evaluation

- Wie effektiv sind die verschiedenen syntaktischen Merkmale?
- Evaluierung von Arabisch-zu-Englisch Übersetzungen
- Testsets von NIST's MT-Eval von 2002 bis 2006, bezeichnet als MT02 bis MT06
- Baseline SMT System: Moses

Evaluation - Genauigkeit der POS Tagger

	POS Accuracy
CM2 (für Development und Test Sets)	94.4%
CRF (Conditional Random Fields)	97.0%
S-POS (simple tagger)	86.8%

Evaluation - Baseline Ergebnisse (BLEU)

	MT04	MT05	MT06
Moses	48.97	53.92	38.40
+ DLM n-gram	49.57	54.42	39.08
Oracle	61.06	66.34	50.11

Evaluation – nach Merkmalen der Parsbäume

	MT04	MT05	MT06
Moses	48.97	53.92	38.40
+ DLM n-gram	49.57	54.42	39.08
+ n-gram + POS	49.47	54.48	39.07
+ n-gram + SEQ-B	49.09	54.11	39.47
+ n-gram + SEQ-C	49.46	54.19	39.07
+ n-gram + CFG	49.53	54.44	39.58
+ n-gram + H	49.44	54.09	33.45

Evaluation - Anteil an Sätzen, die einen Parse haben

	# Sätze	p.p.s%
MT0203	1282287	87.3%
MT04	1075563	81.9%
MT05	744049	82.6%
MT06	1526170	80.7%

Evaluation – Merkmale von POS Taggern und POS- Annotationen des Parsbaums

	MT04	MT05	MT06
DLM n-gram	49.57	54.42	39.08
DLM n-gram + POS	49.47	54.48	39.07
Verbesserung	-0.10	0.06	-0.01
DLM n-gram + CRF	49.74	54.51	39.45
Verbesserung	0.17	0.09	0.37
DLM n-gram + S-POS	49.59	54.60	39.48
Verbesserung	0.02	0.18	0.40

Evaluation – POS Tag

Häufigkeit, Fehlen von POS Typen, Verb Agreement

	MT04	MT05	MT06
+ DLM n-gram	49.57	54.42	39.08
+ S-POS+vn+dn	49.65	54.60	39.67
+ S-POS+allnum	49.65	54.60	39.67
+ S-POS+noall	49.70	54.46	39.69
+ S-POS+verbagr	49.44	54.56	39.55

Evaluation – n-Gramm Precision

Task	System	n-gram Precision (%)			
		1	2	3	4
MT04	n-gram + Syntax	81.86	58.36	41.72	30.28
		81.76	58.48	41.92	30.43
Verbesserung (%)		-0.1%	0.2%	0.5%	0.5%
MT05	n-gram + Syntax	81.31	62.74	47.20	35.54
		83.28	62.96	47.43	35.74
Verbesserung (%)		-0.04%	0.3%	0.5%	0.6%
MT06	n-gram + Syntax	74.43	47.84	31.75	21.50
		74.31	47.92	31.87	21.58
Verbesserung (%)		-0.2%	0.2%	0.4%	0.4%

Fazit

- Nutzen syntaktischer Merkmale
- Nutzen einfacher POS Tagger
- Signifikante Verbesserungen
- Zukünftig: mögliche Nutzung von partiellen Parsern → weniger Informationen als bei Fullparsern, aber mehr als bei POS Taggern

Vielen Dank!

Quellen

- Discriminative Syntactic Reranking for Statistical Machine Translation, S. Carter & C. Monz, AMTA 2010
- <http://www.cs.uni-paderborn.de/fileadmin/Informatik/AG-Kleine-Buening/files/ws11/ml11/folien-neuronale-netze.pdf>
- http://www.cs.jhu.edu/~zfli/pubs/discriminative_lm_for_smt_zhifei_amta_08.pdf