

Übungsblatt 6

1. Grammatik mit Merkmalsconstraints

Erweitern Sie die Grammatik von Folie 7 der letzten Vorlesung so, dass sie, wie die Grammatik vom Ende des Foliensatzes 5, auch einfache Relativsätze behandeln kann. Tun Sie dies in zwei Schritten:

- Ergänzen Sie in der neuen Grammatik CFG-Regeln zur Erzeugung von Relativsätzen.
- Ergänzen Sie die Merkmalsconstraints zu den einzelnen Regeln entsprechend.
- Demonstrieren Sie das Funktionieren der neuen Grammatik, indem Sie den folgenden Satz mit seiner zugehörigen Merkmalsstruktur ableiten:

Der Student studiert das Fach, das er wählte.

2. Evaluation

Die Folien zur Vorlesung zeigen eine Beispielevaluation für eine binäre Klassifikationsaufgabe (Folie 35ff.), die sich auf die Klasse Adj konzentriert.

- Berechnen Sie Precision, Recall und F-Score für die Klasse NAdj.
- Vergleichen Sie die Ergebnisse für NAdj mit den Ergebnissen für Adj. Wie gut ist der Klassifikator Ihrer Meinung nach tatsächlich, um Adjektive zu erkennen?

3. Statistische Modellierung

Wir trainieren den statistischen Klassifikator aus den Vorlesungsfolien auf einigen tausend Sätzen aus dem TIGER-Korpus und erhalten folgende Frequenztafel:

Selbst Artikel?	Nächstes Wort kapitalisiert?		Adj	NAdj
Falsch	Falsch		66	10533
Falsch	Wahr		787	4966
Wahr	Falsch		0	565
Wahr	Wahr		0	1162

- Wie viele Klassen gibt es, wie heißen sie?
- Wie viele Features/Merkmale gibt es? Wie viele Werte haben die Merkmale jeweils? Wie groß ist der Ereignisraum/Menge der Merkmalsmuster?

- c. Für welchen Anteil des gesamten Ereignisraumes hat der Klassifikator Trainingsinstanzen gesehen? Was bedeutet das für die Abdeckung des Klassifikators auf neuen Daten?
- d. Beschreiben Sie die Bedeutung des Modells als Menge von Regeln, die den einzelnen Ereignissen Klassen zuordnen. Wie denken Sie über die Nützlichkeit des Klassifikators?
- e. Welches Grundproblem hat die in der Vorlesung vorgestellte naive statistische Klassifikation, wenn manche Klassen sehr viel kleiner (weniger frequent) sind als andere Klassen?