

5. Übung zur Vorlesung “NLP – Analyse des Wissensrohstoffes Text” im Sommersemester 2008

Dr. Andreas Hotho, Dipl.-Inform. Dominik Benz, Wi.-Inf. Beate Krause

11. Juni 2008

1 Probabilistic Context Free Grammars (PCFG)

1.1 Grundlagen

1. Grenzen Sie den Vorgang des Parsing von PoS-Tagging ab. Welche sprachlichen Eigenschaften lassen sich mit Hilfe von Parsing analysieren?
2. Entwickeln Sie eine probabilistische kontextfreie Grammatik, welche für den Satz “People saw the man on the hill” mindestens zwei (möglichst linguistisch sinnvolle) Analysen liefert. Nutzen Sie die Chomsky Normalform.

1.2 Wahrscheinlichkeiten eines Satzes

1. Welche Algorithmen lassen sich verwenden, um die Wahrscheinlichkeit für einen Satz w_{1m} gegeben die Grammatik G zu errechnen?
2. Berechnen Sie mit den Wahrscheinlichkeiten der S.384 Tabelle 11.2 die Outside Wahrscheinlichkeiten für den Satz “astronomers saw stars with ears”.
 - a) Erstellen Sie eine passende Tabelle.
 - b) Beginnen Sie mit der Startwahrscheinlichkeit im rechten, oberen Feld.
 - c) Arbeiten Sie sich diagonal weiter vor.
 - d) Errechnen Sie $P(w_{1m}|G)$.
3. Wie kann der Inside Algorithmus verändert werden, um den wahrscheinlichsten Parsbaum für einen gegebenen Satz zu finden? Welche Werte würden sich in der Tabelle 11.3 ändern und wie?

1.3 Trainieren einer PCFG

1. Welche Informationen liegen für das Trainieren von PCFGs vor? Was genau umfasst der Prozess des Trainierens?

2. Benutzen Sie wieder das Beispiel aus dem Buch und die Outside Ergebnisse aus der vorherigen Aufgabe, um die Wahrscheinlichkeiten für die Regeln aus der ersten Spalte der Tabelle 11.2 neu zu bestimmen. Wenden Sie dabei den Inside-Outside Algorithmus an. Eine Iteration genügt.

2 Praxisübung - Parsing (Abgabe: Mittwoch, 25. Juni 2008)

Das NLTK bietet verschiedene Parser. In dieser Übung sollen die vorhandenen Parser genutzt und erweitert werden.

1. Schreiben Sie ein Program, welches Sätze mit Hilfe des Recursive Descendent Parser und des Shift Reduce Parsers aus dem NLTK parsen kann.
2. Implementieren Sie selbst einen probabilistischen Left-Corner Parser.
3. Vergleichen Sie die Parser. Verwenden Sie dabei verschiedene Sätze und protokollieren Sie kurz, welche Hauptunterschiede Ihnen auffallen.