

Musterlösungen zum 5. Übungsblatt

Sebastian Padó

12. Dezember 2006

Aufgabe 5.1a

Konfusionsmatrix für NAdj:

	Echte NAdj	Echte Adj
Als NAdj klass.	900	10
Als Adj klass.	80	10

Evaluationsmaße für NAdj:

- Recall: $900 / (900+80) = 900/980 = 91.8\%$
- Precision: $900 / (900+10) = 900/910 = 98.9\%$
- F-Score: $(2*0.918*0.989)/(0.918 + 0.989) = 0.952$

Aufgabe 5.1b

Die Ergebnisse für NAdj sind fast perfekt, sehr viel besser als die Ergebnisse für Adj.

Aufgabe 5.1c

Die Adj-Evaluation. Die Klasse Adj ist die “interessante” Klasse, und die Ergebnisse für diese Klasse sind sehr schlecht. Das Modell weist fast immer NAdj zu, und diese Klasse ist daher sehr gut.

Aufgabe 5.2

- Es gibt zwei Klassen: Adj, NAdj
- Zwei Features, mit je zwei Werten. Die Größe des Ereignisraumes ist das Produkt der Wertanzahl aller Features, hier $2 \times 2 = 4$ Ereignisse.
- Wir haben alle Ereignisse einmal gesehen (100%). Das heisst, für jedes Ereignis auf neuen Daten können wir eine Entscheidung treffen (ob sie richtig ist oder nicht, ist eine ganz andere Frage).
- Für das erste Ereignis (selbst nicht Artikel, nächstes Wort nicht kapitalisiert) weist der Klassifikator die Klasse NAdj zu. Dasselbe tut er für alle anderen Ereignisse – die Klasse Adj wird nie zugewiesen. Das heisst, der Klassifikator ist völlig unbrauchbar.
- Für jedes Ereignis wird die zugewiesene Klasse danach bestimmt, welche am häufigsten mit diesem Ereignis gesehen wurde. Wenn es sehr seltene und sehr häufige Klassen gibt, ist es für die kleinere Klasse sehr schwer, überhaupt für irgendein Ereignis häufiger vorzukommen als die häufige Klasse.