

2. Schriftliche Prüfung zur Übung

Statistische Methoden in der maschinellen Sprachverarbeitung

WS 2013/14

Helmut Schmid

Aufgabe 1) Nennen Sie drei Anwendungen von N-Gram-Sprachmodellen in der Computerlinguistik. (2 Punkte)

Aufgabe 2) In der Vorlesung wurde argumentiert, dass für einen Text T und eine Sprache L gilt:

$$\arg \max_L p(L|T) = \arg \max_L \frac{p(T|L)p(L)}{p(T)} \quad (1)$$

$$= \arg \max_L p(T|L)p(L) \quad (2)$$

$$= \arg \max_L p(T|L) \quad (3)$$

Geben Sie die Begründung für jeden der drei Schritte an. Bei welchem Schritt ist eine zusätzliche Annahme notwendig. Wie lautet sie? (3 Punkte)

Aufgabe 3) Erläutern Sie, wie der Viterbi-Algorithmus bei Hidden-Markow-Modellen 2. Ordnung (=Trigramm-Modellen) arbeitet und geben Sie an, wie die Viterbi-Wahrscheinlichkeiten definiert sind. (3 Punkte)

Aufgabe 4) Begründen Sie, warum der Viterbi-Algorithmus beim Wortart-Taggen mehr Zeit braucht, wenn die Zahl der möglichen Tags eines Wortes oder die Ordnung des HMMs (= Zahl der relevanten vorhergehenden Tags) erhöht wird.

Steigt die benötigte Rechenzeit schneller bei Erhöhung der Zahl der Tags oder bei Erhöhung der Ordnung des HMMs? (3 Punkte)

Aufgabe 5) Warum wird beim PCFG-Parsen die Grammatik oft markowisiert? Welches Problem soll dadurch gelöst werden? (3 Punkte)

Aufgabe 6) Wie ist bei einem linearen Modell die bedingte Wahrscheinlichkeit einer Analyse y für ein gegebenes x definiert? (3 Punkte)

Aufgabe 7) Angenommen Sie haben ein lineares Modell mit 10 verschiedenen Merkmalsfunktionen, von denen jedoch einige nutzlos sind. Wie können Sie die beste Kombination von Merkmalen ermitteln? Was benötigen Sie dafür? (Stichwort: Optimierung von Metaparametern) (3 Punkte)

Aufgabe 8) Mit dem Perzeptron-Algorithmus können die Parameter eines linearen Modelles trainiert werden:

```
for  $t = 1 \dots T$ 
  for  $i = 1 \dots n$ 
     $z \leftarrow \arg \max_{u \in \text{Gen}(x_i)} w \cdot \phi(x_i, u)$ 
    if  $z \neq y_i$ 
       $w \leftarrow w + [\phi(x_i, y_i) - \phi(x_i, z)]$ 
return  $w_s$ 
```

Implementieren Sie den Perzeptron-Algorithmus als Funktion, welche die Zahl der Merkmale, die Zahl der Iterationen T und die Menge der Trainingsdaten (x_i, y_i) in geeigneten Datenstrukturen übergeben bekommt. Sie können annehmen, dass (1) eine Funktion “Gen(x)” existiert, welche für ein gegebenes x die Menge der möglichen Klassen y liefert, und dass (2) eine Funktion “phi(x,y)” existiert, welche die Merkmale für das Objekt x und die Klasse y berechnet und in einer geeigneten Datenstruktur zurückgibt.

Ihr Algorithmus soll den trainierten Gewichtsvektor zurückgeben.

Wenn Sie wollen, können Sie annehmen, dass die Objekte x und die Klassen y einfach als ganze Zahlen repräsentiert werden.

Als Programmiersprache können Sie Perl, Python, C++, oder Java verwenden.

(10 Punkte)

(30 Punkte insgesamt)

Viel Erfolg!