Schriftliche Wiederholungsprüfung zur Vorlesung
Statistische Methoden in der maschinellen Sprachverarbeitung
WS 2017/18
Dozent: Helmut Schmid

Aufgabe 1) Wie lautet das **Theorem von Bayes**? (1 Punkt)

Aufgabe 2) Wozu dient eine **Parameter-Glättung**? Wie unterscheidet sich die Berechnung der NGramm-"Häufigkeiten" für Backoff-Ngramme bei der Kneser-Ney-Glättung und der normalen Backoff-Glättung? (2 Punkte)

Aufgabe 3) Geben Sie die allgemeine Formel an, mit der bei einem Hidden-Markow-Modell 2. Ordnung (d.h. einem Trigramm-HMM) die Wahrscheinlichkeit $p(w_1^n, t_1^n)$ einer Wortfolge w_1^n mit der Tagfolge t_1^n berechnet wird. Was ist bzgl. Satzanfang und Satzende zu beachten?

Welche Wahrscheinlichkeiten müssen konkret für die Wortfolge "Es regnet" und die Tagfolge "PPER VVFIN" multipliziert werden? (3 Punkte)

Aufgabe 4) Warum sind unbekannte Wörter beim Wortart-Taggen problematisch? Wie kann ein HMM-Tagger mit unbekannten Wörtern umgehen? Wie müssen Sie hierfür die "HMM-Formel" anpassen? (3 Punkte)

Aufgabe 5) Die Formel für die Berechnung des χ^2-Testes lautet:

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Wofür stehen die Variablen O_{ij} und E_{ij}? Welche Anwendung des χ^2-Testes haben wir in der Vorlesung betrachtet? Was ist eine Kontingenztabelle und wie sieht sie aus? (3 Punkte)

Aufgabe 6) Wie funktioniert die **Addiere-1**-Glättung (mit Formel) und warum ist sie nicht geeignet, um Wortwahrscheinlichkeiten zu glätten? Bei welchen Voraussetzungen ist die Addiere-1-Glättung optimal? (2 Punkte)

Aufgabe 7) Erklären Sie das **EM-Training** am Beispiel des unüberwachten Trainings von Wortart-Taggern. Welche Schritte umfasst das Verfahren? Was sollte gegeben sein? Mit welchem Algorithmus kann das EM-Training bei einem HMM-Tagger effizient implementiert werden? (3 Punkte)

Aufgabe 8) Angenommen Sie trainieren einen HMM-Tagger mit dem Forward-Backward-Algorithmus. Wie können Sie die **erwartete Häufigkeit** (= Aposteriori-Wahrscheinlichkeit) des Tags t an der Position des Wortes w_k aus den Forward-Wahrscheinlichkeiten $\alpha_t(k)$ und Backward-Wahrscheinlichkeiten $\beta_t(k)$ berechnen?

Wie berechnen Sie die erwartete Häufigkeit des Tagpaares t und t' an den Positionen der Wörter w_k und w_{k+1}? (2 Punkte)

>>>>>>>>>>>>>>> weiter auf der nächsten Seite >>>>>>>>>>>>>>>>

Aufgabe 9) Wie werden **Precision**, **Recall** und **F-Score** bei der Evaluierung eines Parsers berechnet. Welche Häufigkeiten brauchen Sie für die Berechnung? (2 Punkte)

Aufgabe 10) Beschreiben Sie ausführlich alle Schritte zur Durchführung eines Evaluierungsexperimentes bei einem Wortart-Tagger.

Erklären Sie außerdem, wie man den **Vorzeichentest** (Binomialtest) anwendet, um zu berechnen, ob der Tagger signifikant besser als ein anderer (Baseline-)Tagger ist.
(3 Punkte)

Aufgabe 11) Lineare Modelle berechnen für ein Objekt x und eine Klasse y einen Merkmalsvektor $f(x, y)$ und multiplizieren ihn mit einem Gewichtsvektor θ. Diejenige Klasse, bei der das Produkt am größten ist, wird ausgegeben.

Solche Modelle können mit dem **Perzeptron**-Algorithmus trainiert werden. Geben Sie Pseudocode für das Perzeptron-Training an. (2 Punkte)

Aufgabe 12) Wie kann auf Basis des Produktes $f(x, y) \cdot \theta$ von Merkmalsvektor $f(x, y)$ und Gewichtsvektor θ eine **Wahrscheinlichkeitsverteilung** $p(y|x)$ (loglineares Modell) definiert werden? (2 Punkte)

Aufgabe 13) CRF-Tagger können mit dem Gradientenanstiegs-Verfahren trainiert werden.

- Welche Zielfunktion wird beim Training eines CRF-Taggers maximiert?
- Die Ableitung dieser Zielfunktion ist die Differenz zwischen der beobachteten Häufigkeit der Merkmale im Korpus und der erwarteten Häufigkeit der Merkmale. Mit welchem Algorithmus berechnet man diese erwarteten Häufigkeiten?
- Welche Aufgabe hat die Lernrate?
- Wozu dient eine Regularisierung? Welche Methoden der Regularisierung werden bei CRFs verwendet? (2 Punkte)

(30 Punkte insgesamt)

Viel Erfolg!