

Анализ текстов на естественных языках

Контрольная работа

Задание 1. Дан язык $L = \{aa, bb, abab, baba, aaaa, bbbb\}$ над алфавитом $\Sigma = \{a, b\}$. Вычислить оценки максимального правдоподобия для параметров биграммной языковой модели со сглаживанием по схеме “прибавление единицы”, с их помощью вычислить вероятность строки $abba$.

Задание 2. Решается задача аннотирования гласных и согласных в словах переменной длины над алфавитом $\Sigma = \{a, b, c\}$, набор помет $T = \{C, V\}$. Обучающая выборка $(\mathcal{X}, \mathcal{Y}) = \{(a, V), (b, C), (ab, VC), (ba, CV), (abb, VCC), (abba, VCCV)\}$. Построить биграммную скрытую марковскую модель, параметры модели задать на основе оценок максимального правдоподобия. Найти наиболее вероятную аннотацию для строки $bbaabb$ в соответствии с построенной моделью.

Задание 3. Требуется построить систему автоматической расстановки переносов для русского языка на основе скрытой марковской модели. Выпишите спецификацию модели.

Задание 4. Дана вероятностная контекстно-свободная грамматика $\langle N, \Sigma, R, S, q \rangle$, где $N = \{A, B, C, S\}$, $\Sigma = \{a, b\}$, $R = \{S \rightarrow CB \mid BC, C \rightarrow BA \mid AB, A \rightarrow a, B \rightarrow b\}$, $q(S \rightarrow CB) = 0.7, q(S \rightarrow BC) = 0.3, q(C \rightarrow BA) = 0.4, q(C \rightarrow AB) = 0.6, q(A \rightarrow a) = 1.0, q(B \rightarrow b) = 1.0$. Найти наилучший (наиболее вероятный) разбор для строки abb .