Lehrstuhl Informatik 6
RWTH Aachen University
Prof. Dr.-Ing. H. Ney

Aachen, 15. 06. 2016

R. Schlüter,
K. Irie,
A. Zeyer

Automatic Speech Recognition

4. Exercise

In this exercise we will use word graphs for confidence estimation and Bayes risk decoding.
Please use Python to solve the programming tasks.
IMPORTANT: Please send your solution code to `zeyer@i6.informatik.rwth-aachen.de` with a copy (cc.) to `irie@i6.informatik.rwth-aachen.de` .
Submission Deadline: 29. 06. 2016 at the beginning of the exercise session.

Task 4.1 Word graph, confidence estimation and Bayes risk decoding

This exercise is based on the materials of the exercise 3. Please refer to the corresponding sheet for the description of the materials.

(a) Implement the forward-backward algorithm to compute the sum over all paths through each edge of the word graph. You can possibly adapt your code from the exercise 3. (8 P)

(b) Use (a) to compute the word posterior probabilities of all word graph edges, for each of the provided word graphs. (4 P)

(c) Use (b) to derive time frame word posteriors and compute the confidence of the best recognized word sequence from the word graph by summing over the time frame word posteriors for all frames of the corresponding word graph edge. (4 P)

(d) Use the time frame word posteriors from (c) to rescore the word graph. Extract the best hypotheses after rescoring and report the corresponding word error rate. Compare it to the word error rate obtained in the exercise 3 with the standard decision rule (4 P)

Task 4.2 Minimum Bayes Risk with Hamming Distance and Constant Length Word Sequences

Assume a simplified acoustic model without alignment problem, with acoustic observations on a per word position level. The joint path probability for a word sequence $W = w_1^N$ and the corresponding observation vector sequence $X = x_1^N$ with a bigram language model then becomes:

$$p(X, W) = \prod_{n=1}^{n} p(w_n|w_{n-1})p(w_n|x_n)$$

Furthermore assume word sequences of constant length N only. Finally consider the following Hamming loss function to be used in Bayes decision rule:

$$L(v_1^N, w_1^N) = \sum_{n=1}^{N} (1 - \delta_{v_n, w_n})$$

and substitute it into the corresponding posterior risk needed for Bayes decision rule:

$$R(w_1^N, x_1^N) = \sum_{v_1^N} p(v_1^N | x_1^N) L(v_1^N, w_1^N)$$

(a) Rewrite the posterior risk in terms of suitable word posteriors and provide the definition of the word posteriors. (7 P)

(b) Define the corresponding Bayes decision rule based on your result from a). (3 P)

(c) **Bonus problem:** Derive an efficient algorithm to compute the word posteriors. (10 P)