

(520|600).666

Information Extraction from Speech and Text

Homework # 5

Due March 10, 2009.

Review Chapter 5 from *Statistical Methods for Speech Recognition* by Frederick Jelinek.

1. **Dynamic Time Warping:** This is the popular name of a technique that was successfully used for small-vocabulary, *isolated-word* recognition for a considerable period of time. Many variants of DTW have been proposed, and the following is an illustrative example.

For each word v in the vocabulary \mathcal{V} , a *template* $\mathbf{B}(v) = \langle b_1(v) b_2(v) \dots b_{l(v)}(v) \rangle$ and, for each b_j , a *cost* $c(a|b_j)$ over the acoustic symbols $a \in \mathcal{A}$ is provided. In order to recognize an utterance $\mathbf{A} = a_1, a_2, \dots, a_k$, its *best cost*

$$C_{\mathbf{A}}(v) = \min_{j_1, \dots, j_k} \sum_{t=1}^k c(a_t | b_{j_t}), \quad 1 = j_1 \leq j_2 \leq \dots \leq j_k = l(v), \quad (1)$$

is computed for each word in the vocabulary, and the word with the lowest cost is chosen. Some additional restrictions are (usually) placed on the possible sequences j_1, \dots, j_k over which the minimum is computed, such as $j_{t+1} \leq j_t + 2$ for all $t = 1, 2, \dots, k - 1$.

- (a) Explain the name *dynamic time warping*. (Hint: set $k = 8$ and $l(v) = 5$, and pick any two sequences j_1, \dots, j_8 that satisfy the conditions above. For each sequence, plot the points $\{(t, j_t), t = 1, \dots, 8\}$ on a rectangular grid with $l(v)$ horizontal lines and k vertical lines, and “connect the dots.” Examine the two resulting mappings from the interval $[1, k]$ to the interval $[1, l(v)]$.)
- (b) Over roughly how many admissible sequences j_1, \dots, j_k does the minimum of (1) need to be evaluated?
- (c) Design an algorithm to efficiently compute the minimum of (1). (Hint: In a manner similar to the Viterbi algorithm, think about two different sequence-prefixes j_1, \dots, j_l and j'_1, \dots, j'_l with $(l, j_l) = (l', j'_l)$ as two paths that end up at the same “grid node.” How do their completions $j_1, \dots, j_l, j_{l+1}, \dots, j_k$ and $j'_1, \dots, j'_l, j_{l+1}, \dots, j_k$ compare in the minimization?)
- (d) How is the DTW model connected with fenonic baseforms? In particular, can you suggest a procedure for estimating the costs $c(a|b_j)$ given the template and sample utterances of each word?

2. **Coarse-to-Fine Viterbi Search:** Replace the Viterbi search (pp 22-23) for finding the most likely state sequence s_1, \dots, s_k , given the observations y_1, \dots, y_k , and initial state s_0 , by a succession of Viterbi searches on *increasingly complex* HMMs as follows.

- (a) Let $\mathcal{S} = \{1, \dots, c\}$ denote the set of HMM states, and \mathcal{Y} the output alphabet. Assume that output probabilities $q(y|s' \rightarrow s)$ are associated with the arcs $s' \rightarrow s$ of the HMM.

Hierarchically cluster the HMM states into an approximately balanced binary tree, the $\approx \frac{c}{2}$ first-level clusters comprising of 2 states each, which are then clustered into $\approx \frac{c}{4}$ second-level clusters of 4 states each, and so on. This clustering can be arbitrary.

Let \mathcal{S}_0 and \mathcal{S}_1 denote the top split of the tree, containing complementary subset of \mathcal{S} . Let \mathcal{S}_{00} and \mathcal{S}_{01} denote the split of \mathcal{S}_0 , \mathcal{S}_{10} and \mathcal{S}_{11} denote the split of \mathcal{S}_1 , and so on.

- (b) Construct a 2-state HMM with *super-states* $\mathcal{S}^{(2)} = \{0, 1\}$. Designate the initial state of this HMM to be 0 if, in the original HMM, $s_0 \in \mathcal{S}_0$, and designate it to be 1 otherwise.

Construct arcs $i \rightarrow j$ with output “probabilities”

$$q^{(2)}(y|i \rightarrow j) = \max_{s' \in \mathcal{S}_i, s \in \mathcal{S}_j} q(y|s' \rightarrow s) \quad i, j \in \{0, 1\}, y \in \mathcal{Y},$$

omitting arcs $i \rightarrow j$ whenever there are *no arcs* from any state $s' \in \mathcal{S}_i$ to a state $s \in \mathcal{S}_j$ in the original HMM.

Construct the k -stage trellis for this HMM, with exactly 2 states per trellis stage (time step).

- (c) Use the Viterbi algorithm to find the most likely path $s_1^{(2)}, s_2^{(2)}, \dots, s_k^{(2)}$ through this trellis.
- (d) Along (only) this winning path in the trellis, “shatter” or *refine* each winning *super-state* into two smaller *super-states* based on the hierarchical clustering of \mathcal{S} . Note that in some trellis stages, state $0 \in \mathcal{S}^{(2)}$ will be shattered to obtain states 00 and 01 based on the division of \mathcal{S}_0 into \mathcal{S}_{00} and \mathcal{S}_{01} , while for other stages, state $1 \in \mathcal{S}^{(2)}$ will be shattered to obtain states 10 and 11 based on the division of \mathcal{S}_1 into \mathcal{S}_{10} and \mathcal{S}_{11} . Each trellis stage will have exactly 3 states drawn from the state-space $\mathcal{S}^{(3)} = \{00, 01, 1\} \cup \{0, 10, 11\}$.

- (e) Construct arcs $i \rightarrow j$ for states in the refined trellis, with output “probabilities”

$$q^{(3)}(y|i \rightarrow j) = \max_{s' \in \mathcal{S}_i, s \in \mathcal{S}_j} q(y|s' \rightarrow s) \quad i, j \in \{0, 1, 00, 01, 10, 11\}, y \in \mathcal{Y},$$

unless there are *no arcs* from any state $s' \in \mathcal{S}_i$ to any state $s \in \mathcal{S}_j$ in the original HMM.

Use the Viterbi algorithm to find the most likely path $s_1^{(3)}, s_2^{(3)}, \dots, s_k^{(3)}$ in the refined trellis.

- (f) Iterate the trellis refinement of Step (d) and Viterbi decoding of Step (e) until the winning path *cannot be refined further*, i.e. it consists only of individual states in the original HMM.

Answer the following questions for this iterated Viterbi search procedure.

- (i) After Step (c), if, say, $s_t^{(2)} = 0$ for some t , is there any guarantee that the Viterbi path s_1, \dots, s_k in the original HMM passes through some state $s_t \in \mathcal{S}_0$? Why or why not?
- (ii) Argue why the procedure described above finds the correct Viterbi path in the original HMM.
- (iii) What is the best-case and worst-case computational complexity of this iterated Viterbi procedure? Compare it with the regular Viterbi search, which has complexity $O(c^2k)$.
- (iv) Discuss how the hierarchical clustering of the states ought to be done, so that performance closer to the best-case than the worst-case is likely to be obtained.

Continue working on Project #2.