HMM Theory and Practice

STUDENTS-HUB.com

Training & Recognition

- Major advantage of HMMs is the availability of a 'toolkit' of powerful, well-founded mathematical methods for HMM manipulation
- The **Baum-Welch** algorithm is used to train the parameters of a set of HMMs given a set of training data
- Viterbi Decoding is used to classify an unknown speech pattern in terms of the sequence of HMMs which is most likely to have produced it

The Recognition Problem

• Given a sequence of acoustic feature vectors

$$Y = \{y_1, \dots, y_T\}$$

we want to find the sequence of words

$$W = \{w_1, ..., w_L\}$$

such that the probability

is maximized.

• If $M = \{M_1, ..., M_K\}$ is the sequence of HMMs which represents W, then P(W | Y) = P(M | Y)

STUDENTS-HUB.com

Bayes' Theorem

 Computation of the probability P(M | Y) is made possible using Bayes' Theorem

$$P(W \mid Y) = \frac{p(Y \mid W)P(W)}{p(Y)}$$

- *P(W)* is the "language model probability"
- *p(Y | W)* is the "acoustic model probability"
- Bayes Theorem has been referred to as the "fundamental theorem of speech recognition"!

The Baum-Welch Algorithm

- The Baum-Welch algorithm is the method which is normally used for HMM parameter estimation
- Given a set of HMMs M₀ and a set of speech patterns Y, Baum's theorem defines how to produce a new model set M₁ such that

 $P(Y \mid M_1) \ge P(Y \mid M_0)$

- Baum-Welch algorithm applies this method repeatedly until a HMM M_n is found which (locally) maximizes $P(Y \mid M)$
- Baum's theorem only valid for particular classes of state output PDF

Notes on B-W Reestimation

- The Baum-Welch algorithm is only guaranteed to find a locally optimal HMM set hence choice of M₀ can be important
- Baum-Welch is a **supervised** training algorithm which requires labelled speech data
- The labelling need **not** be at the same level as the HMM set phoneme level HMMs can be trained using data labelled orthographically at the phrase or sentence level
- For large applications B-W reestimation can be **very** computationally expensive

- Viterbi Decoding is the algorithm which is used to find the sequence of HMMs which is most likely to have generated a given speech pattern
- Based on **Dynamic Programming**
- Viterbi Decoding illustrates the type of computation typically done with HMMs

Viterbi Decoding (1)



Q: How can *M* have generated *Y*?A: Via a state sequence of length T



Function of State Sequence



STUDENTS-HUB.com

Viterbi Decoding (2)

Construction of 'state-time trellis'



Constructing the State-Time Trellis

- Simple Rule:
 - Connect node (*i*,*t*) of the trellis to node
 (*j*,*t*+1) if and only if there is a transition
 between state *i* and state *j* in the HMM with
 probability *a_{ij}* greater than zero

Basic Probability Calculation



Viterbi Decoding (3)

- Let $X = \{x_1, \dots, x_T\}$ be a state sequence of length T
- The joint probability of *Y* and *X* is given by:

$$p(Y,X) = b_{x_1}(y_1) \prod_{t=2}^T a_{x_{t-1}x_t} b_{x_t}(y_t)$$

- i.e. the product of the state-output and state transition probabilities along the state sequence
- The <u>optimal</u> state sequence is the sequence X such that p(Y,X) is maximized
- *p(Y)* is the sum of *P(Y,X)* over all sequences *X*

Viterbi Decoding (4)



STUDENTS-HUB.com

Isolated Speech Recognition



STUDENTS-HUB.com

Connected Speech Recognition

New transitions connect end of every model to $\mathbf{y}_1 \quad \mathbf{y}_2 \quad \mathbf{y}_3 \quad \mathbf{y}_4 \quad \mathbf{y}_5$ Y_{t-1} Y_t Ут start of every model Ο Ο Ο Ο Ο Ο Ο \cap

Connected Speech Recognition



STUDENTS-HUB.com

Further explanation of Viterbi decoding

STUDENTS-HUB.com



STUDENTS-HUB.com



STUDENTS-HUB.com



STUDENTS-HUB.com



STUDENTS-HUB.com



STUDENTS-HUB.com



STUDENTS-HUB.com





STUDENTS-HUB.com



STUDENTS-HUB.com



STUDENTS-HUB.com



STUDENTS-HUB.com



STUDENTS-HUB.com



STUDENTS-HUB.com

Trace-back



Trace-back



Trace-back

