# Level Building with Hidden Markov Models

Asmaa Rabie

Cairo University

*asmaa_rabie@eng.cu.edu.eg*

January 1, 2016

# Why Level Building

Given a sequence of observations to be matched against HMM models requires either:

- Using a segmental solution, so that the whole gesture is segmented into smaller gestemes that can can be matched against a specific model, however there are several setbacks:
  - Its hard to find the exact segmentation points,
  - The number of gestemes constituting a gesture is unknown
  - Exhaustive search for all possibilities is exponential $M^L$, where $M$ denotes the number of gestemes and $L$ denotes the number of gestemes per gesture
- Using a non-segmental approach like the level building algorithm to save up the segmentation process

# Input

- $L$: is the max number of gestemes per gesture
- $M$: is the number of gesteme HMMs, each model is specified by $\lambda(A, B, \pi)$, where:
  - $A$: is the transition matrix where $a_i j$ is the probability of being in state $j$ given that the previous state was $i$. $1 \leq i, j \leq N$
  - $N$: is the number of states of this models
  - $B$: is the emission matrix, where $b_j(O_t)$ is the probability of seeing $O$ at time $t$ given that the current state is $j$, $1 \leq t \leq T$
  - $T$: is the number of observations
  - $\pi$: is the initial probability where $p_i$ is the probability of starting at state $i$

# Level Building Algorithm (LBA)

For each level $L$ and each model $q$ do a Viterbi match to find:

- $P(l, t, q)$: The level $l$'s output best probability up to each time frame $t$ from each model $q$, where $1 \leq l \leq L$, $1 \leq t \leq T$ and $1 \leq q \leq M$
- $\hat{P}(l, t) = \max_{q}[P(l, t, q)]$: the model that produced this best probability at each frame $t$
- $B(l, t, q)$: The backpointer at each time frame $t$ from each model $q$
- $\hat{B}(l, t) = B(l, t, \underset{q}{argmax} P(l, t, q))$: the level output backpointer $t$
- $\hat{W}(l, t) = \underset{q}{argmax} P(l, t, q)$: the level output gesteme indicator (i.e. the best model $q$ that generated this best probability)

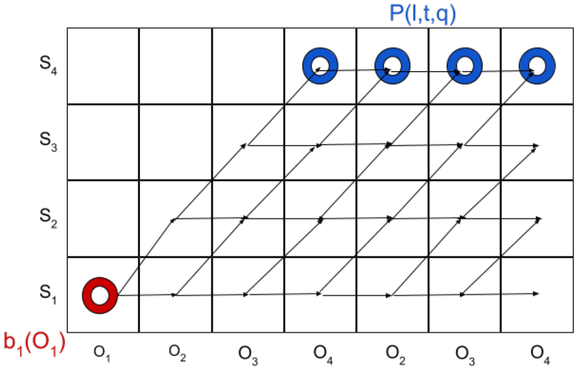
# Level Building Algorithm (LBA)

## at level 1

1. Initialization
   $\delta_t(j)$ is the joint probability of partial state $j$ at frame $t$
   - $\delta_1(1) = b_1^q(O_1)$
   - $\delta_1(j) = 0, j = 2, 3, ..., N$
2. Recursion
   for $2 \leq t \leq T, 1 \leq j \leq N$
   $$\delta_t(j) = \max_{j-1 \leq i \leq j} [\delta_{t-1}(i) * a_{ij}^q] * [b_j^q(O_t)]$$
3. Termination
   - $P(l, t, q) = \delta_t(N)$
   - $B(l, t, q) = 0$

# Level Building Algorithm (LBA)

## at higher levels



# Level Building Algorithm (LBA)

## at higher levels

1. Initialization

   Pick up initialization values from previous output

   - $\delta_1(1) = 0$
   - $\delta_t(1) = max[\hat{P}(l-1, t-1), a_{ij}^q * \delta_{t-1}(1)] * [b_1^q(O_t)]$

   $\alpha_t(j)$ is the back pointer array, which indicates the time frame that lead to state $j$ at time frame $t$

$$\alpha_t(1) = \begin{cases} t-1 & if\, \hat{P}(l-1, t-1) > a_{ij}^q * \delta_{t-1}(1) \\ \alpha_t - 1(1) & otherwise \end{cases}$$

2. Recursion

   for $2 \leq t \leq T, 1 \leq j \leq N$

   $\delta_t(j) = \max\limits_{j-1 \leq i \leq j} [\delta_{t-1}(i) * a_{ij}^q] * [b_j^q(O_t)]$

   $\alpha_t(j) = \alpha_{t-1}[argmax \delta_{t-1}(i) * a_{ij}^q]$

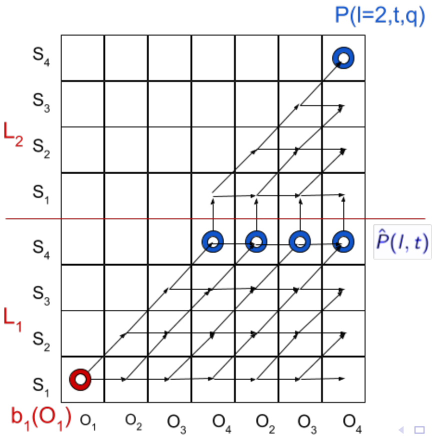
# Level Building Algorithm (LBA)

## at higher levels

3. Termination
    - $P(l, t, q) = \delta_t(N)$
    - $B(l, t, q) = \alpha_t(N)$

# Level Building Algorithm (LBA)

## at higher levels



# References

L. Rabiner, S.E. Levinson, *A speaker-independent, syntax-directed, connected word recognition system based on hidden Markov models and level building*. IEEE Transactions on Acoustics, Speech and Signal Processing, 1985.

# The End