GAUSSIAN MIXTURE MERGING IN KOREAN SPEECH RECOGNITION SYSTEMS

John Morgan
United States Military Academy
Department Of Foreign Languages
745 Brewerton Road
West Point, NY 10996
Email: john.morgan@usma.edu
Phone: 845-938-2966
Fax: 845-938-3585

ARL Sponsor: Melissa Holland Ph.D.
U.S. Army Research Laboratory
Adelphi, Maryland

November 4, 2003

CTELL ARL Symposium 2003

THE PROBLEM

- recognize language learners
- tolerate non-native speech
- consider proficiency
- separate recognition from evaluation

HIDDEN MARKOV MODELS

- states: $Q = \{q_1 q_2 \ldots q_N\}$
- transition probability matrix: A_{ij}
- observation likelihoods: $b_i(o_t)$

markovgen.eps

Figure 1: Markov Model

OBSERVATION LIKELIHOODS

- Simple Observation Likelihood for state j:

$$b_j(o_t) = \mathcal{N}(o_t; \mu_j, \Sigma_j) \tag{1}$$

- $\mathcal{N}(\cdot; \mu, \Sigma)$ multivariate Gaussian
- μ mean vector
- Σ covariance matrix

MIXTURES

- **Theorem 1.** *A Linear combination of independent normally distributed random variables is normally distributed*

- observation likelihood for state j:

$$b_j(o_t) = \sum_{m=1}^{M} c_{jm} \mathcal{N}(o_t; \mu_{jm}, \Sigma_{jm}) \tag{2}$$

- M number of mixture components

- c_{jm} weight of m'th mixture

MERGING

- English observation likelihood from state i

$$e_i = \sum_{m=1}^{M} c_{im} \mathcal{N}(o_t; \mu_{im}, \Sigma_{im})$$

- Korean observation likelihood from state j

$$k_j = \sum_{l=1}^{L} d_{jl} \mathcal{N}(o_t; \mu_{jl}, \Sigma_{jl})$$

- $$\begin{cases} e_i \oplus k_j(o_t) & = & \sum_{n=1}^{M+L} g_n \mathcal{N}(o_t; \mu_n, \Sigma_n) \\ g_i & = & u \frac{c_i}{\sum_{n=1}^{M+L} c_n} \\ u & = & w \text{ if Korean mixture} \\ u & = & 1 - w \text{ otherwise} \end{cases}$$

- u language weight

CONFUSION MATRIX

- make phone state map
- Korean forced alignment
- English phone-loop alignment
- make confusion matrix with results
- normalize matrix entries
- use matrix entries as mixture weights