IDIAP

Martigny - Valais - Suisse

morris@idiap.ch
http://www.idiap.ch/

Two variants in HMM ASR with soft missing data - outline of theory and initial experiments

-

1. Combining FC experts, MD theory and EM for ML expert weights estimation

2. ASR with soft missing data, where uncertainty is based on data utility as well as reliability

Combining FC experts, MD theory and EM for ML expert weights estimation

-

Problem. For model $p(x|k) = \sum_l P(l|k)p(x|k, l)$ **with fixed pdfs** $p(x|k, l)$, where (k, l) are class & mix comp. index, find weights $P(l|k)$ to maximise $p(X|W)$.

EM solution

E-step: Find expected log lik. in terms of new/old params

$$Q^{new} = -\sum_{n,k} P^{old}(k|x_n)\ln(p^{new}(x_n, k)) \quad \text{k = class index} \quad (1)$$

M-step: Find new params to maximise Q

$$P^{old}(k|x_n) = \frac{p^{old}(x_n|k)P^{old}(k)}{\sum_{k'} p^{old}(x_n|k')P^{old}(k')} \quad \text{Bayes' rule} \quad (2)$$

$$P^{new}(k) = \frac{1}{N}\sum_n P^{old}(k|x_n) \quad \text{diff } Q \text{ \& eq. to zero} \quad (3)$$

For one mix pdf per class, simply replace (k) by (j, k) above.

*All probabilities above are **estimated** probabilities for a **given set** of fixed model parameters,* Θ.

Replace (j) above by (j, k, l) where (l) is indicator index for the set of MUAE events *"substream combination* (l) *is clean and its complement is noisy or missing"*.

(k, l) are class & substream expert index

Problem. For x **noisy,** $p(x|j, k, l)$ **not** $=$ $p(x|j, k, l, \Theta)$

Solution. Replace $p(x|j, k, l, \Theta)$ **by** $E[p(x|j, k, l, \Theta)]$

$$p(x|j, k, l, \Theta) \cong E[p(x|j, k, l, \Theta)] \quad \text{min. var. estimate} \qquad (4)$$

$$= p(x_l|j, k, l, \Theta)E[p(x_l'|x_l, j, k, l, \Theta)] \qquad \text{factorise} \qquad (5)$$

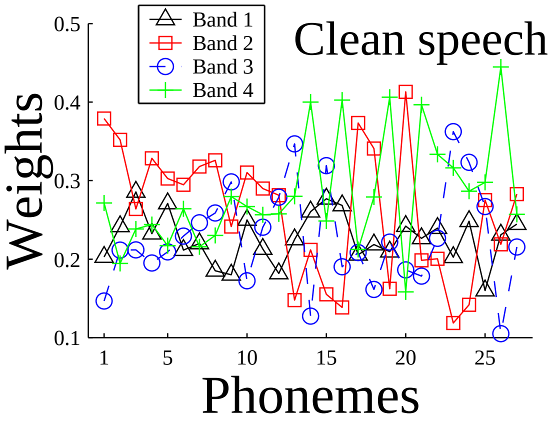
$$p(x_l'|x_l, j, k, l, \Theta) = p(x_l'|j, k, l, \Theta) \quad \text{diag. cov. Gausn} \qquad (6)$$

$$p(x_l|j, k, l, \Theta) \qquad \text{easily evaluated Gaussian marginal} \qquad (7)$$

$$E[p(x_l'|j, k, l, \Theta)] = \alpha_{j, k, l} \quad \text{easily evaluated constant} \qquad (8)$$

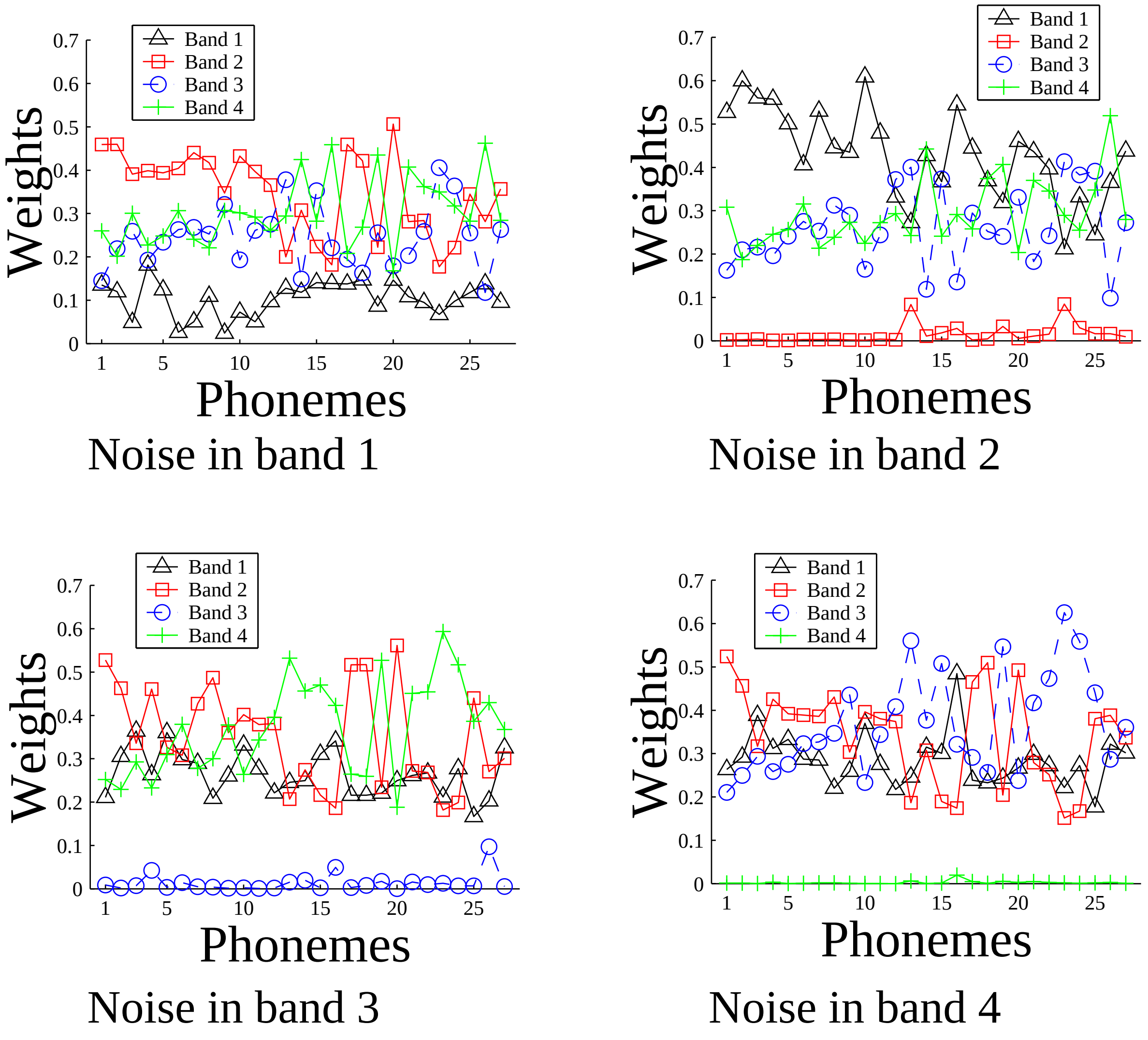
Step (3) requires diag. covariance, which applies only at level of mix components. If required, weights /expert/class can be obtained from weights /expert/class/mix comp. by summation:

$$P(k, l) = \sum_j P(j, k, l).$$



Clean speech

Fig. left shows ML wts for each phoneme for each of 4 subband experts, with clean speech.



Noise in band 1

Noise in band 2

Noise in band 3

Noise in band 4

Figs. above show ML weights $P(l|k)$, for each phoneme for each of 4 subband experts, for artificial noise in bands 1..4.

Initial results for ML FC weighting

-

Initial N95 tests diverge from the above model in two ways:

- weight estimation assumed $p(x|j, k, l) \propto p(x_l|j, k, l, \Theta)$, i.e. $E[p(x_l'|j, k, l, \Theta)] = \alpha_{j, k, l}$ was constant over j,k,l.
- weights were used with phoneme posteriors combination, though they were derived for likelihood combination.

noise	Equal weights	ML weights
clean	24.3	26.0
band 1	44.0	38.3
band 2	27.3	27.3
band 3	35.4	30.2
band 4	31.9	28.7
siren	33.9	33.9

Table 1:
WER for band limited noises at SNR 0,
using MFCCs & 4 subband experts only

While multiband ASR shows strong advantage over baseline only with band limited noise, the ML weighting described here applies equally to multistream, automatically detecting which stream combination experts give the most benefit.

ASR with soft missing data, where uncertainty is based on data utility as well as reliability

-

Normal MAP decoding uses $W = \arg max_W P(W|X)$, where

$$P(W|X) \propto P(W)p(X|W) \cong P(W)\prod_t p(x_t|q_t) \qquad (1)$$

Problem. For x noisy, $p(x|q)$ **not =** $p(x|q, \Theta)$

Solution. Replace $P(W|X)$ by $E = E[P(W|X)]$

$$E = P(W)E\left[\frac{p(X|W)}{p(X)}\right] = P(W)\int\frac{p(X|W)p(X|X^o)}{p(X)}dX \qquad (2)$$

$$= \prod_t P(q_t|q_{t-1})A_t \text{ where } A_t = \int\frac{p(x_t|q_t)}{p(x_t)}p(x_t|x_t^o)dx_t \qquad (3)$$

ASR with hard MD uses $x = (x^o, x^u)$, giving

$$A = \int_0^{x^o}\frac{p(x^o|q)p(x^u|x^o, q)}{p(x^o)p(x^u|x^o)}\,\frac{p(x^o)p(x^u|x^o)}{\int_0^{x^o}p(x^u|x^o)dx^u}dx^u \qquad (4)$$

$$A \propto p(x^o|q)\int_0^{x^o}p(x^u|x^o, q)dx^u = A^o A^u \qquad (5)$$

ASR with soft MD decision

Several methods have been suggested (last year) for better reflecting uncertainly in deciding which data is missing.

e.g. If $P(missing) = \alpha$, could express A_i, for each GM component, as $A_i = \alpha A_i^u + (1 - \alpha) A_i^o$, instead of $A_i^o A_i^u$.

ASR with soft MD

A more direct approach is to model the observation data pdf $p(x_i | x_i^o) = s(x_i)$ as a pdf.

This is feasible only if we assume

$$\frac{p(x|q)}{p(x)} p(x|x^o) = \prod_i \frac{p(x_i|q)}{p(x_i)} p(x_i|x_i^o) \qquad (1)$$

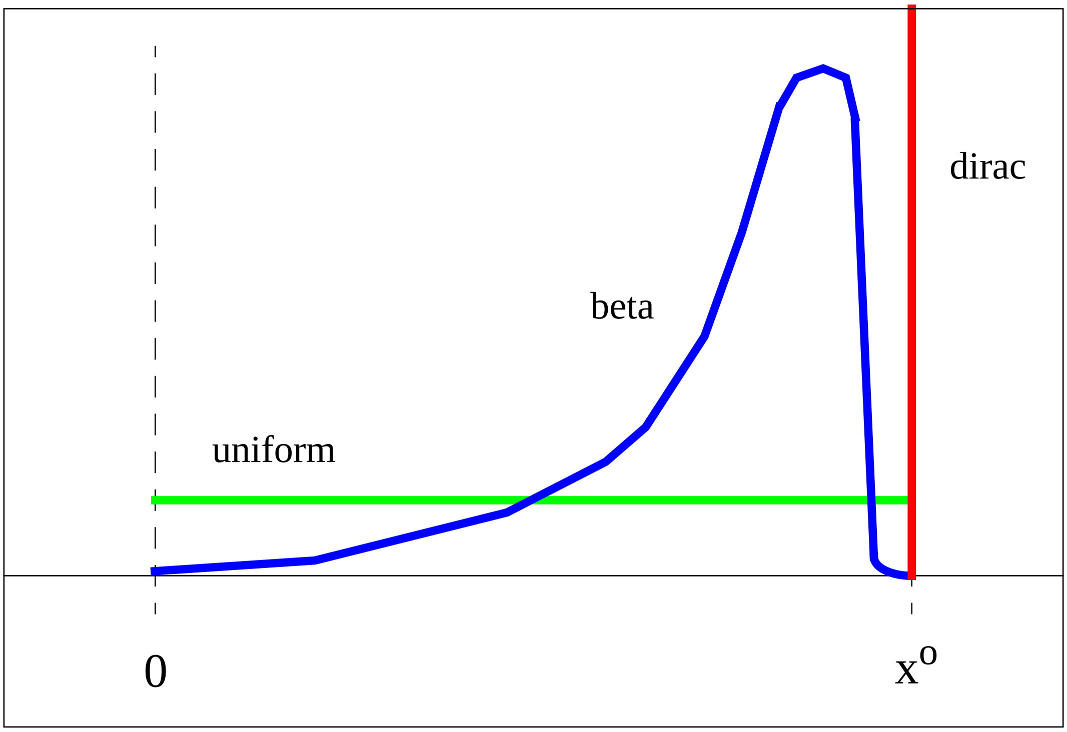
and tabulate $f(x, k) = \dfrac{p(x|q_k)}{p(x)}$

'A' can then be evaluated simply as $A = \prod_i \sum_x f_k(x) s(x) dx$

In this way every coefficient of observation data is preprocessed to become a pdf instead of single value.

- Pdf used for “missing data” with “bounds constraint” = clean data pdf, restricted to $[0, x^o]$.

- Present “soft data” approach **treats all data as uncertain**.



Pdf for each data point varies between dirac pdf for “certain” data, to uniform pdf for totally “missing” data

- data weight is determined by pdf width

- permits weighting not only of reliability, but also utility.