WEIGHTED PRINCIPAL COMPONENT MLLR FOR SPEAKER ADAPTATION

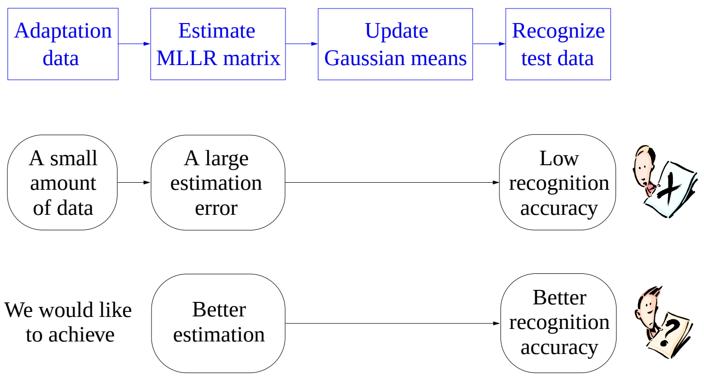
Sam-Joo Doh and Richard M. Stern

Department of Electrical and Computer Engineering and School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213, USA
{sjdoh, rms}@cs.cmu.edu, http://www.cs.cmu.edu/~robust/

ABSTRACT

We present two new speaker adaptation methods which apply principal component analysis to maximum likelihood linear regression (MLLR) framework. If we apply MLLR after transforming the baseline mean vectors by their eigenvectors, the variance of the estimates for the MLLR matrix are inversely proportional to their corresponding eigenvalues. We describe two techniques to reduce the variance of the estimation, Principal Component MLLR (PC-MLLR) and Weighted Principal Component MLLR (WPC-MLLR). In experiments using sentences from the 1994 DARPA Wall Street Journal evaluation, the use of WPC-MLLR provided a relative reduction in word error rates of 15.1% for non-native speakers and 6.0% for native speakers compared to conventional MLLR.

The procedure of MLLR speaker adaptation



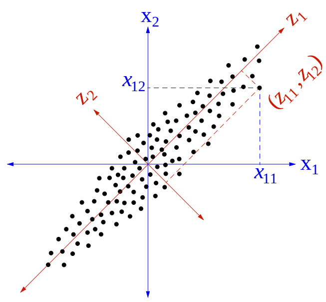
Related Work

- **Block diagonal MLLR**: Reduce the number of parameters for MLLR matrix [1]
- **Principal Component Analysis (PCA)**
  - Feature selection [5]
  - "Eigenvoices": represent speaker variation [3]
  - Correlation between phoneme classes [2]
- Previous PCA work was performed primarily for dimensionality reduction and is not directly related to MLLR.

Principal Component Regression

(*Example*)

Linear regression: $(x_{i1}, x_{i2}) \rightarrow (y_i)$



- **In the original domain:** $y = Xc + \varepsilon$

$$\begin{bmatrix} y_1 \\ y_2 \\ : \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ : & : \\ x_{n1} & x_{n2} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ : \\ \varepsilon_n \end{bmatrix}$$

- We can transform X and c to Z and γ by using eigenvector matrix (V_X) and eigenvalue matrix (Λ_X).

$$(X^T X) \cdot V_X = V_X \cdot \Lambda_X$$

$$Z = X \cdot V_X, \text{ and } \gamma = V_X^T \cdot c$$

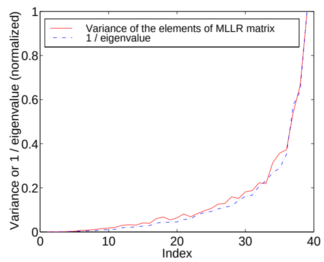
- **In the eigendomain:** $y = Z\gamma + \varepsilon$

$$\begin{bmatrix} y_1 \\ y_2 \\ : \\ y_n \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \\ : & : \\ z_{n1} & z_{n2} \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ : \\ e_n \end{bmatrix}$$

- In the eigendomain, **the variance of $\hat{\gamma}_j$ is inversely proportional to the corresponding eigenvalue of** $(X^T X)$ [4]
  - In this example, (variance of $\hat{\gamma}_1$) << (variance of $\hat{\gamma}_2$)
- We can select the **top *p* principal components** of the estimates $\hat{\gamma}_j$ to reduce the overall variance of the estimates.
  - Eliminate highly variable components
  - Obtain the estimate $\hat{c} = V_{(p)} \cdot \hat{\gamma}_{(p)}$ using top p components

Principal Component MLLR (PC-MLLR)

- MLLR assumes $\hat{\mu}_m = A\mu_m + b$.
  - μ_m and A correspond to X and c in the above example.
- We can apply Principal Component Regression to MLLR by using eigenvetors and eigenvalues obtained from Gaussian means and variances.
- Comparison of the **average variances** of the elements of the MLLR multiplication matrix in the eigendomain obtained from actual data and the **inverse of eigenvalues** confirms that **they are indeed the same** (in figure).



Weighted Principal Component MLLR (WPC-MLLR)

- Instead of eliminating non-principal components of the estimates of the MLLR matrix, we **use all the components after weighting**.
- Find weights for the estimate $\hat{\gamma}_j$ to minimize the mean square error.

$$\hat{\gamma}'_j = \omega_j \cdot \hat{\gamma}_j \rightarrow \text{minimize } E(\hat{\gamma}'_j - \gamma_j)^2$$

$$\omega_j = \frac{\lambda_{Xj}}{\lambda_{Xj} + \sigma_\varepsilon^2/\gamma_j^2} \approx \frac{\lambda_{Xj}}{\lambda_{Xj} + k}$$

- **Large eigenvalue** λ_{Xj} → **small variance** of $\hat{\gamma}_j$ → **large weight** $\omega_j \approx 1$
- **Small eigenvalue** λ_{Xj} → **large variance** of $\hat{\gamma}_j$ → **small weight** $\omega_j \approx 0$

- **WPC-MLLR adaptation steps**

(0) Pre-calculate eigenvectors, eigenvalues, and weights
(1) Transform baseline mean vectors by their eigenvectors
(2) Estimate MLLR multiplication matrix (row vector $\hat{\alpha}_r$) and shift vector b in eigendomain using adaptation data
(3) Multiply each MLLR matrix element by its weight: $\hat{\alpha}'_{rj} = \omega_j \cdot \hat{\alpha}_{rj}$
(4) Re-calculate the shift vector b using $\hat{\alpha}'_r$
(5) Transform $\hat{\alpha}'_r$ back to the multiplication matrix A
(6) Adapt the baseline mean vectors using $\hat{\mu}_m = A\mu_m + b$

EXPERIMENTS

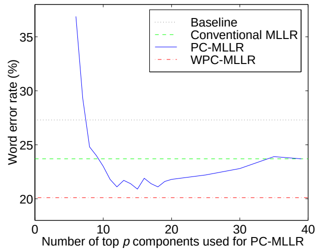
- Speech recognition system: SPHINX-III (continuous HMMs)
- Test data: DARPA 1994 Wall Street Journal evaluation
  - Spoke 3 (non-native speakers) and Spoke 0 (native speakers).
- Adaptation data: 5 sentences for each speaker (supervised mode)

- WPC-MLLR provides relative improvements in recognition accuracy compared to conventional MLLR of 15.1% for non-native speakers and 6.0% for native speakers.

Word error rates
for selected data from the 1994 WSJ evaluation after adaptation
(Relative improvement over the baseline is shown in parenthesis)

Adaptation Method	s3-94 data (Non-native)	s0-94 data (Native)
Baseline (unadapted)	27.3%	21.9%
Conventional MLLR	23.7% (13.1%)	18.3% (16.4%)
PC-MLLR	20.9% (23.4%)	18.0% (17.8%)
WPC-MLLR	20.1% (26.3%)	17.2% (21.4%)

- In PC-MLLR, if the number of components is too small, WER increases rapidly. If all the components are used, WER asymptotes to conventional MLLR.



Word error rates for each adaptation method for s3-94 data
as a function of the **number of principal components** used for PC-MLLR

SUMMARY

- **Principal Component Analysis (PCA)** has been applied to the MLLR framework for speaker adaptation
- The **variance of the estimate** for linear regression in the eigendomain is **inversely proportional** to corresponding **eigenvalue**.
- **Principal Component MLLR (PC-MLLR)**
  - Eliminates highly variable components and selects the top *p* principal components
  - Reduces the variance of the estimates and improves speech recognition accuracy
- **Weighted Principal Component MLLR (WPC-MLLR)**
  - Uses all the components after weighting to minimize the mean square error
  - Gives greater weight to components with larger eigenvalues
  - Achieves better recognition accuracy than PC-MLLR and conventional MLLR

ACKNOWLEGEMENT

This research was sponsored by the Space and Naval Warfare Systems Center, San Diego, under Grant No. N66001-99-1-8905. The content of the information in this publication does not necessarily reflect the position or the policy of the US Government, and no official endorsement should be inferred. Sam-Joo Doh is partially supported by Korea Telecom Overseas Education Program.

REFERENCES

[1] M.J.F. Gales, D. Pye and P. C. Woodland, "Variance Compensation Within the MLLR Framework for Robust Speech Recognition and Speaker Adaptation," *Proc. of ICSLP*, pp.1832-1835, 1996.

[2] Z. Hu, *Understanding and adapting to speaker variability using correlation-based principal analysis*, Ph.D. Thesis, Oregon Graduate Institute of Science and Technology, 1999.

[3] R. Kuhn *et al*., "Eigenvoices for Speaker Adaptation," *Proc. of ICSLP*, pp.1771-1774, 1998.

[4] R. H. Myers, *Classical And Modern Regression With Applications*, PWS-KENT Publishing Company, Boston, 1990.

[5] J. Nouza, "Feature Selection Methods for Hidden Markov Model-Based Speech Recognition," *Proc. of the 13th International Conf. on Pattern Recognition*, Vol.2, pp.186-190, 1996.