Optimizing PLLR features for Spoken Language Recognition

M. Diez, A. Varona, M. Penagarikano, L.J. Rodriguez-Fuentes, G. Bordel

GTTS, Department of Electricity and Electronics, University of the Basque Country UPV/EHU, Spain

Abstract

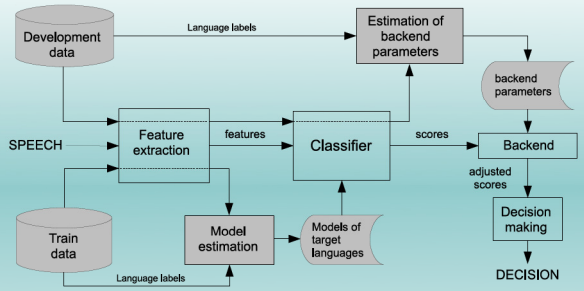
Phone Log-Likelihood Ratios (PLLRs): recently introduced features for spoken language and speaker recognition systems
Effective way of retrieving acoustic-phonotactic information into frame-level vectors

In this work:
•Extend the search of reduced representations of PLLRs
•Evaluate the effect of using larger temporal contexts: Shifted Delta transformation is applied on reduced sets of PCA-projected PLLRs

The Task

Spoken language recognition is a pattern recognition task that consists of recognising the language spoken in an utterance by computational means

General structure of a SLR system:



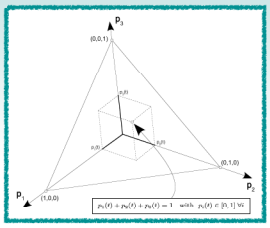
PLLR Features

- Phone decoder: N phone units, S states per model
- $p(i|s,t)$ Acoustic posterior probability of each state s $(1 \le s \le S)$ of each phone model i $(1 \le i \le N)$ at each frame t, directly provided by the phone decoder
- Acoustic posterior probability of a phone unit i at each frame t:

$$p(i|t) = \sum_{\forall s} p(i|s,t)$$

- Assuming a classification task with flat priors, and taking the log-likelihood ratios, the obtained distributions are nearly Gaussian:

$$LLR(i|t) - \log \frac{p(i|t)}{\frac{1}{(N-1)}(1-p(i|t))} \qquad i - 1, ..., N$$



- Brno University of Technology phone decoder for Hungarian (61 phonetic units, each featuring a three-state model). Three non-phonetic units integrated into a single non-phonetic unit: 59 PLLR features computed at each frame
- Feature vector augmented with first order deltas: 118-dimensional feature vector

iVector Approach

- Under the iVector modeling assumption, an utterance GMM supervector (stacking GMM mean vectors) is defined as:

$$M = m + Tw$$

M	utterance dependent GMM mean supervector
m	utterance independent mean supervector
T	total variability matrix
w	iVector

- The iVector approach maps high-dimensional input data to a low-dimensional feature vector, retaining most of the relevant information

Generative modeling approach is applied in the i-vector features space, the distribution of i-vectors of each language being modeled by a single Gaussian distribution. Scores are computed as follows:

$$score(f,l) = N(w_f; \mu_l, \Sigma)$$

Data and evaluation measures

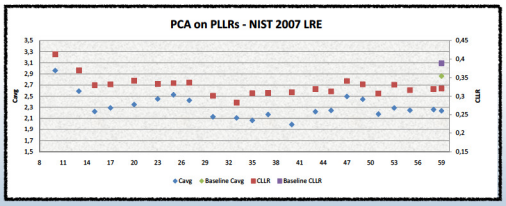
NIST 2007 LRE: conversational speech across telephone channels, 14 target languages.

NIST 2011 LRE: pairwise language detection task, 24 target languages.

Systems compared in terms of:

(1)The average cost performance C_{avg}
(2)The Log-Likelihood Ratio Cost C_{LLR}
(3)The primary measure C_{avg}^{24} for NIST 2011 LRE, which averages the actual cost for the 24 pairs with the highest minimum cost

Dimensionality reduction

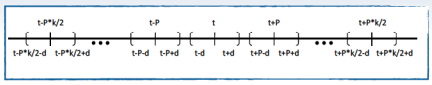


Baseline system (without PCA):
2.86 C_{avg}
0.390 C_{LLR}.

Performance significantly enhanced by PCA projection → decorrelation of the feature space

Shifted Deltas on PLLRs

SD-PLLR features are specified by four parameters, N-d-P-k:



Search for the optimal configuration:

PCA Dim		C_{avg}	C_{LLR}
13	PLLR	2.59	0.370
13	SD- PLLR 13-2-3-7	1.71	0.260
15	PLLR	2.23	0.330
15	SD- PLLR 15-2-3-7	1.94	0.264
17	PLLR	2.29	0.332
17	SD- PLLR 17-2-3-7	1.73	0.241

SD-PLLR configuration	C_{avg}	C_{LLR}
13-2-1-7	2.39	0.346
13-2-2-7	1.91	0.279
13-2-3-7	1.71	0.260
13-2-4-7	2.02	0.297
13-2-5-7	2.46	0.347

SD-PLLR configuration	C_{avg}	C_{LLR}
13-1-3-7	2.04	0.286
13-2-3-7	1.71	0.260
13-3-3-7	2.03	0.277

Results on NIST 2011 LRE

System	C_{avg}	C_{LLR}	$\%C_{avg}^{24}$
Baseline	5.18	0.982	12.12
SD-PLLR 13-2-3-7	4.10	0.826	10.48

Conclusions

•**Projection of the features enhances performance** of the system, due in part to the decorrelation of the feature space achieved by applying PCA

•Best results are achieved by projecting PLLRs into **33 dimensions**: a 26% relative improvement in terms of C_{avg} w.r.t. the baseline system

•**Shifted-Delta** over PCA-projected PLLR features, reaches 1.73 and 4.10 C_{avg} for the NIST 2007 and 2011 LRE datasets, **40% and 21% relative improvements** with regard to using the PLLR features, respectively

ICPR 22nd INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
IAPR IEEE computer society
24-28 August 2014 Stockholm, Sweden

Universidad del País Vasco Euskal Herriko Unibertsitatea
GTTS Tecnologías Software