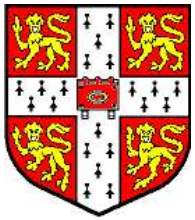
Diarisation Research at CUED

Sue Tranter and Srinivasan Umesh

1st May 2004



Cambridge University Engineering Department

Diarisation Research

- Cluster Voting
  - Definition
  - Description of Implementation
  - Current Results
  - Future Work
- LIMSI-Style Segmentation
  - Motivation
  - Preliminary Results

Cluster Voting - Definition

Aim:

- Combine information from different diarisation systems to improve diarisation performance.

Combination after final output allows:

- Different architectures of systems to be combined directly.
- Different stopping criteria to be used.
  (Stopping criteria can be the most important factor in diarisation systems.)

Difficulties:

- Relatively few 'tokens' (unlike e.g. ROVER for STT).
- Potential complexity issues when large differences between inputs.

Cluster Voting - Implementation

Stage 1: Generate CVOS

- Find all the members of the Cluster Voting Output Set (CVOS) which minimise the sum of the diarisation error rate (DER) from the output to the 2 inputs.
- Currently done with exhaustive search of all possible segment clusterings after several techniques to reduce the complexity.

Stage 2: Pick Final Output from CVOS

- Could be rule-based e.g. 'always combine if $<$ x duration'
- Could be confidence score-based for more than 2 inputs.
- Current 2-input methods based on BIC-type model selection criterion.

Cluster Voting Results

bneval03 data (6 shows, LDC forced alignment, no collar, 0.3s smoothing)

System	TOTAL DER	#shows DER $\leq$ inputs	#shows DER $\geq$ inputs
CUED-diary-bic (input 1)	25.12	-	-
CUED-diary-cost (input 2)	27.09	-	-
Best CVOS score	22.79	6	0
Worst CVOS score	29.44	0	6
standard-BIC-judge, full cov*	*24.62*	2	2
standard-BIC-judge, diag cov*	*23.90*	3	1
standard-BIC-judge, 128mix GMM*	*23.76*	5	1
IDIAP-BIC-judge, full cov	*25.02*	2	2
IDIAP-BIC-judge, 15mix GMM†	*23.48*	5	0

* best result when varying α value. † best result when varying number of Gaussians in base GMM.Breakdown by show is given as an appendix.

Best system gives 1.64% absolute reduction in DER over the best input.

Cluster Voting - Future Work

Documentation:

S. E. Tranter
Cluster Voting for Speaker Diarisation
Tech Report CUED/F-INFENG/TR-476, Cambridge Uni. Engineering Dept

Generalisation:

- Reduce complexity by using spkr-mapping between inputs to restrict possible output clusterings.
- Test on different systems, architectures and data sets.
- Allow more than 2 inputs.

LIMSI-Style Segmentation - Motivation

bneval03 data (6 shows, LDC forced alignment, no collar, 0.3s smoothing)

	Num Segs	Ideal DER	Actual DER	WER †
LIMSI-feb04-seg	571	7.07	18.95	10.22
LIMSI-feb04-spkr	"	"	12.06	10.18
CUED-rt03s-stt	876	9.10	(58.25)	10.65
CUED-dec03-spkr	869	9.09	25.12	(10.76)

† WER using CUED 10xRT system, generated April 2004

• LIMSI segmenter output and spkr output both outperform CUED system for diarisation and STT

Reference: **The LIMSI Broadcast News transcription system**

Jean-Luc Gauvain, Lori Lamel and Gilles Adda

Speech Communication, Volume 37, Issues 1-2, May 2002, Pages 89-108

LIMSI-Style Segmentation - Progress

Preliminary Implementation:

- Use existing CUED classifier to get WB/NB/music labels
- Use dual-phone recogniser to remove long silences
- Use LIMSI-style segmentation/clustering algorithm [NB - parameters have been chosen on the test data]

bneval03 data (6 shows, LDC forced alignment, no collar, 0.3s smoothing)

	Num Segs	Ideal DER	Actual DER	WER
CUED-rt03s-stt	876	9.10	(58.25)	10.65
CUED-dec03-spkr	869	9.09	25.12	(10.76)
CUED-LIMSI-style	718	8.13	29.73	10.44
+ P1 gender/silence	641	7.60	29.16	-
+ CUED-dec03-spkr	"	"	22.81	(10.56)

- Still very preliminary work, but some reduction in WER and DER.

Appendix - Cluster Voting Results by Show

System	ABC	VOA	PRI	NBC	CNN	MNB	TOTAL
CUED-diary-bic (input 1)	32.03	20.78	21.40	32.06	37.92	**10.74**	25.12
CUED-diary-cost (input 2)	29.26	19.82	20.48	31.56	**37.18**	29.34	27.09
Best CVOS score	26.71	18.43	18.11	29.84	37.18	10.74	22.79
Worst CVOS score	34.58	22.48	23.56	33.78	37.92	29.34	29.44
BIC-judge, full cov*	30.30	19.94	*19.15*	32.06	37.92	**10.74**	*24.62*
BIC-judge, diag cov*	30.30	19.94	***18.11***	*31.05*	37.92	**10.74**	*23.90*
BIC-judge, 128mix GMM*	*27.66*	20.78	*19.15*	*30.83*	**37.18**	**10.74**	*23.76*
IDIAP-BIC, full cov	32.85	20.48	21.15	32.06	**37.18**	**10.74**	*25.02*
IDIAP-BIC, 15mix GMM†	30.30	*19.27*	***18.11***	***29.84***	**37.18**	**10.74**	*23.48*

* best result when varying α value

† best result when varying number of Gaussians in base GMM