Advances in Structural Metadata for RT-04 at CUED

M. Tomalin and P.C. Woodland

9th November 2004

Cambridge University

Overview

- From RT-03f to RT-04.
- CTS and BN Slash Unit Boundary Detection (SUBD) systems.
- CTS Filler Word Detection (FWD) systems.
- CTS Interruption Point Detection (IPD) systems.
- Work in Progress.
- Future Plans.

From RT-03f to RT-04

Structural Metadata Extraction (SMD) tasks attempted for RT-03f:

- CTS SUBD

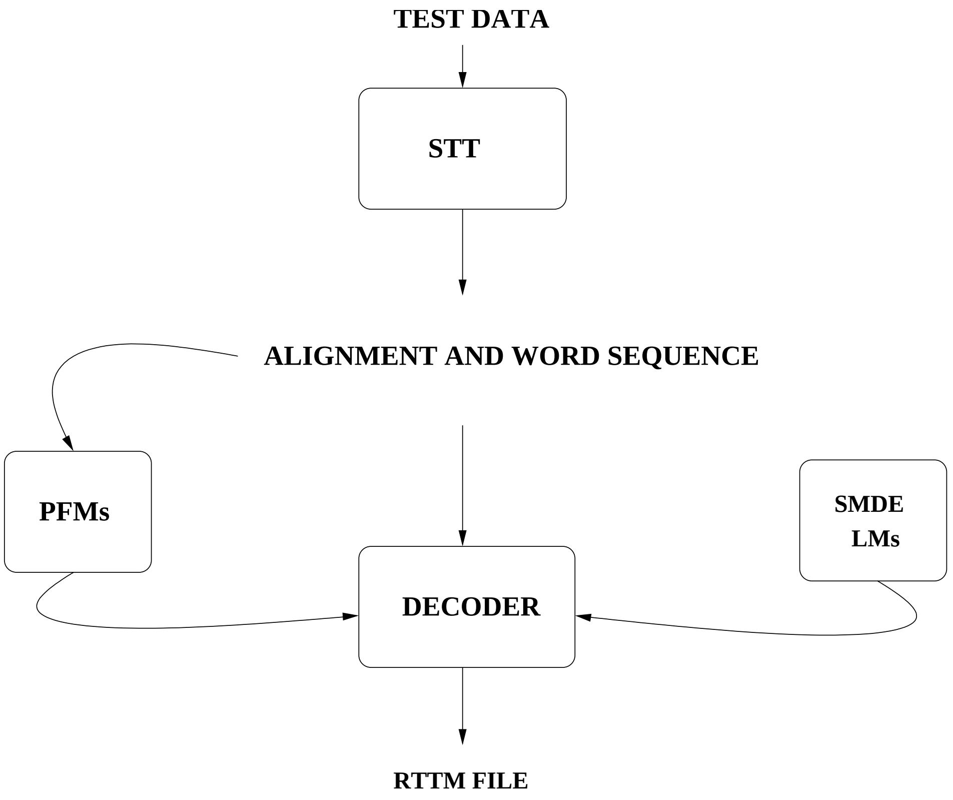
SMD tasks attempted for RT-04:

- CTS SUBD
- BN SUBD
- CTS FWD
- CTS IPD

Three of the CUED SMD systems were built for RT-04.

General System Architecture

The SMD systems used same generic architecture:



General System Architecture

CTS SMD systems:

- input: audio files, CUED 20xRT CTS STT output.[1]
- task-specific Language Models (LMs).
- task-specific Prosodic Feature Models (PFMs).
- 1-Best lattice-based Viterbi Decoder.

BN SMD systems:

- input: audio files, CUED 20xRT BN STT output.[2]
- task-specific LMs.
- task-specific PFMs.
- 1-Best lattice-based Viterbi Decoder.

[1] Evermann et al., 'Development of the 2004 CU-HTK English CTS systems', Proc. Fall 2004 RT-04 Workshop

[2] Kim et al., 'Recent Developments at Cambridge in Broadcast News Transcription', Proc. Fall 2004 RT-04 Workshop

Training and Test Data for CTS

The following sets of CTS training data were used:

Name	**ctsrt04**	**ctsrt04_v1.0**	**ctsrt03**
Epoch	2004	2004	2003
Released	07/09/04	04/06/04	2003
Spec	V6.2 (v1.1)	V6.2 (v1.0)	V5
Hours	c.40	c.40	c.30

These training data sets will be referred to collectively as the ‘EARS CTS’ data.

The following CTS dev sets were used:

Name	**ctsdev03**	**ctseval03**	**ctsdev04**
Epoch	2003	2003	2004
Spec	V6.2 (v1.1)	V6.2 (v1.1)	V6.2 (v1.1)
Hours	c.1.5	c.1.5	c.3

Training and Test Data for BN

The following sets of BN training data were used:

Name	**bnrt04**	**bnrt04_v1.0**	**bnrt03**
Epoch	2004	2004	2003
Released	07/09/04	04/06/04	2003
Spec	V6.2 (v1.1)	V6.2 (v1.0)	V5
Hours	c.20	c.20	c.20

These training data sets will be referred to collectively as the 'EARS BN' data.

The following BN dev sets were used:

Name	**bndev03**	**bneval03**	**bndev04**
Epoch	2003	2003	2004
Spec	V6.2 (v1.1)	V6.2 (v1.1)	V6.2 (v1.1)
Hours	c.1.5	c.1.5	c.3

SUBD for CTS

SUBD results using EARS CTS training data:

SYSTEM	**%Err (DEL/INS/ERR)**		
	dev03	**eval03**	**dev04**
PFM (ctsrt04)	33.6/69.4/132.6	35.2/64.9/133.6	30.2/68.2/131.3
PFM+ctsrt04_fg	31.8/15.1/57.9	31.5/14.0/56.8	29.2/15.7/56.2
PFM+ctsrt04_cl40-tg	33.1/20.3/63.9	33.3/18.7/62.6	30.8/19.7/61.9
PFM+ctsrt04_fg+cl40-tg	31.8/14.8/**57.0**	31.3/13.8/**56.1**	29.1/14.7/**54.4**

NB: All results in these slides obtained using mdeval-v17 with the options '-w -W -t 1.00' set.

PFM trained using ctsrt04 data only.

Interpolated SULMs perform better than independent SULMs.

DEL rates c.15% abs higher than INS rates for all dev sets.

Large Training Data Sets for CTS SUBD

Need to overgenerate SUs to reduce DEL rate:

Only c.100 hrs of EARS CTS training data, so c.1800 hrs of STT WordWave (WW) data mapped to approximate the V6.2 SU annotations.

The mapping rules:

- **full-stop → statement SU boundary**
- **comma → statement SU boundary**
- **question mark → question SU boundary**

Word-based and class-based SULMs were built using mapped WW data.

SUBD for CTS

SUBD results using EARS CTS + WW training data:

SYSTEM	**%Err (DEL/INS/ERR)**		
	dev03	**eval03**	**dev04**
PFM+WW_fg	29.9/46.3/91.3	30.5/46.4/91.8	28.8/47.6/91.1
PFM+ctsrt04_fg+cl40-tg	31.8/14.8/57.0	31.3/13.8/56.1	29.1/14.7/54.4
+ WW_fg	30.7/15.4/**56.7**	30.4/14.3/**55.8**	28.1/15.3/**54.2**

PFM trained using ctsrt04 data only.

WW_fg achieves lower DEL rate than interpolated EARS SULMs.

WW_fg and EARS SULMs interpolated: Err falls by c.0.3% abs.

SUBD for BN

SUBD results using EARS BN training data:

SYSTEM	**%Err (DEL/INS/ERR)**		
	dev03	**eval03**	**dev04**
PFM (bnrt04)	45.2/40.2/110.2	47.3/42.2/107.9	52.0/49.1/134.0
PFM+bnrt03_tg	45.8/17.1/66.1	44.9/20.1/68.8	51.7/24.8/79.8
PFM+bnrt04_v1.0_tg	49.7/15.4/68.6	50.2/15.0/68.5	56.7/19.2/79.8
PFM+bnrt04_tg	50.4/16.0/69.9	49.4/17.2/70.2	55.9/19.9/79.0
PFM+bnrt03_cl40-tg	42.5/22.2/68.0	44.3/24.4/72.5	50.7/28.6/82.7
PFM+bnrt04_v1.0_cl40-tg	49.1/17.1/68.3	49.4/21.2/74.6	55.7/23.5/82.2
PFM+bnrt04_cl40-tg	50.2/17.5/69.5	45.2/20.6/69.0	56.1/25.6/84.8
PFM+EARS SULMs	46.1/14.8/**63.4**	45.4/15.3/**63.9**	53.7/21.7/**78.8**

PFM trained using bnrt04 data only.

DEL rates c.30% abs higher than INS rates for all dev sets.

Large Training Data Sets for BN SUBD

Need to overgenerate SUs to reduce DEL rate:

Only c.60 hrs EARS BN training data, so two STT BN data sets mapped:

Name	**db98**	**bn2003**
Epoch	1998	2003
Hours	c.90	c.4000

These data sets were mapped using same rules as WW data.

Word-based and class-based SULMs were built using mapped BN data.

Large Training Data Sets for BN SUBD

SUBD results using EARS BN + mapped BN data:

SYSTEM	**%Err (DEL/INS/ERR)**		
	dev03	**eval03**	**dev04**
PFM+db98_tg	29.6/35.4/67.9	31.4/44.2/80.6	40.9/45.1/89.4
PFM+db98_cl40-tg	28.0/42.9/74.4	30.1/52.7/87.8	39.1/52.6/95.7
PFM+bn2003_cl40-tg	37.1/26.9/67.4	42.4/30.1/76.8	48.4/36.2/88.6
PFM+EARS SULMs	46.1/14.8/63.4	45.4/15.3/63.9	53.7/21.7/78.8
+ db98 SULMs	42.4/16.6/61.7	42.9/16.7/63.1	52.0/22.4/77.9
+ bn2003 SULMs	41.0/17.2/**61.0**	42.1/16.8/**62.5**	51.5/22.8/**77.8**

PFM trained using bnrt04 data only.

Mapped SULMs reduce DEL rate by c.3% abs on average.

Mapped SULMs reduce ERR rate by c.2% abs on average.

FWD for CTS

The FWD systems consisted of:

- Word-based and class-based Filler Word Language Models (FWLMs).
- A Filler Word PFM trained using ctsrt04 data.
- 1-Best lattice-based Viterbi Decoder.

FWD for CTS

FWD results using EARS CTS training data:

SYSTEM	**%Err (DEL/INS/ERR)**		
	dev03	**eval03**	**dev04**
ctsrt03_tg	35.7/12.4/49.0	36.6/12.8/50.1	31.6/9.7/41.6
ctsrt04_tg	30.0/14.8/**45.9**	32.6/16.4/49.8	26.7/11.9/39.0
ctsrt03_cl40-tg	45.5/12.8/59.1	46.3/13.9/60.1	41.5/10.8/52.8
ctsrt04_cl40-tg	41.0/14.3/55.8	41.2/16.6/58.3	36.4/13.6/50.2
fw_interp	31.8/13.8/46.4	33.7/14.6/**49.2**	27.7/10.8/**38.9**
+ PFM (ctsrt04)	33.4/18.8/52.2	36.0/19.2/55.2	30.2/14.1/44.3

fw_interp = interpolated ctsrt03 and ctsrt04 tgs and cl40-tgs.

The ctsrt04 PFM **increases** ERR by c.6% abs on average.

IPD for CTS

The IPD systems consisted of:

- Word-based and class-based Interruption Point Language Models (IPLMs).
- An Interruption Point PFM trained using ctsrt04 data.
- 1-Best lattice-based Viterbi Decoder.

IPD for CTS

IPD results using EARS CTS training data:

SYSTEM	**%Err (DEL/INS/ERR)**		
	dev03	**eval03**	**dev04**
ctsrt03_tg	51.6/12.5/64.2	53.0/11.9/65.0	49.6/11.6/61.2
ctsrt04_tg	45.7/16.0/61.7	48.0/14.8/62.8	43.6/14.7/58.2
ctsrt03_cl40-tg	52.0/19.6/71.6	55.3/22.0/77.3	53.9/22.4/76.3
ctsrt04_cl40-tg	52.9/20.2/73.0	53.2/17.5/70.7	49.6/17.9/67.5
ip_interp	49.3/12.3/61.5	51.3/11.4/62.7	47.1/11.4/58.5
+ PFM (ctsrt04)	45.7/15.7/**61.4**	48.5/13.7/**62.2**	43.9/14.2/**58.1**

ip_interp = interpolated ctsrt03 and ctsrt04 tgs and cl40-tgs.

PFM decreases ERR by c.0.4% abs.

RT-04 Eval Results

Results for CUED SMD RT-04 Evaluation Systems:

SYSTEM	**%Err (ERR only)**			
	dev03	**eval03**	**dev04**	**eval04**
CTS FMD (spch)	52.2	55.2	44.3	45.8
CTS FMD (ref)	25.3	25.4	25.5	27.4
CTS IPD (spch)	61.4	62.2	58.1	63.5
CTS IPD (ref)	42.8	42.1	44.5	47.2
CTS SUBD (spch)	56.7	55.8	54.2	56.5
CTS SUBD (ref)	52.0	50.6	45.2	46.2
BN SUBD (spch)	61.0	62.5	77.8	72.2
BN SUBD (ref)	57.5	60.6	75.1	71.1

CTS eval04 performance in line with dev set performance for all tasks.

dev04 and eval04 sets for BN SUBD harder than dev03 and eval03 sets.

CTS SUBD: from RT-03f to RT-04

For RT-03f, the following CTS SUBD system was constructed:

- LDC V5 training data (c.40 hrs).
- PFM; 10 prosodic features used.
- Interpolated tg, cl40-tg, and fg SULMs.
- Posterior decoding scheme which ignored SU subtype info.

For RT-04, the following CTS SUBD system was constructed:

- LDC V5 and V6.2 training data (c.100 hrs in total).
- Mapped WW SULM training data (c.1500).
- PFM; 10 prosodic features used.
- Interpolated cl40-tg, and fg SULMs.
- Viterbi 1-Best decoding scheme which preserved SU subtype info.

CTS SUBD: from RT-03f to RT-04

Difficult to compare RT-03f and RT-04 system performance:

- RT-03f: SUB errors not scored; V5 MDE annotation spec.
- RT-04: SUB errors scored; V6.2 MDE annotation spec.

Results using V5 and V6.2 versions of the eval03 scoring ref files:

SYSTEM	DEL	INS	SUBS	%Err (DEL/INS)
RT-03f_sys/V5_ref	33.1	19.3	11.7	64.1 (52.4)
RT-03f_sys/V6.2_ref	34.1	21.2	10.9	66.1 (55.2)
RT-04_sys/V5_ref	32.0	15.1	13.9	61.0 (47.1)
RT-04_sys/V6.2_ref	30.4	14.3	11.2	55.8 (44.7)

RT-04 sys ERR rates between c.5% and c.11% abs lower than RT-03f sys ERR rates.

Work In Progress: Interpolation Weights

Interpolation Weights (IWs) for SMD LMs calculated automatically were suboptimal; IMs for RT-04 LMs selected by hand.

Current approach - insert SU tokens only after relevant words in training data:

> < s > OKAY **SU_S** ARE WE READY **SU_Q** I THINK WE SHOULD GIVE **SU_I** OKAY **SU_S** ... < /s >

Alternative approach - insert SU tokens after **every** word in training data:

> < s > OKAY **SU_S** ARE **SU_N** WE **SU_N** READY **SU_Q** I **SU_N** THINK **SU_N** WE **SU_N** SHOULD **SU_N** GIVE **SU_I** OKAY **SU_S** ... < /s >

Alternative approach enables LM prob streams to be calculated automatically...

Work In Progress: Prosodic Feature GMMs

Current Cart-style decision tree PFMs require

- training data to be downsampled.
- PFM probs to be divided by priors.

Preferable to model the data without downsampling/dividing by priors...

Alternative: GMM-based PFMs:

- Use prosodic features that are modelled well using GMMs.
- Obtain prosodic feature vectors for each SMD event subtype from training data.
- Construct GMM for each SMD event subtype.
- Train GMMs using standard tools, increasing mixtures.
- Obtain prob from each SMD event subtype GMM for each feature vector in test data.
- Place GMM probs on arcs of lattice and decode as usual.

Future Plans

Current plans for SMD research include the following:

- complete automated interpolation weight scheme.
- complete GMM-based prosodic feature modelling work.
- improve the performance of the BN SUBD system.
- explore the interactions between the various SMD tasks.

References

For more information about CUED RT-04 SMD systems:

Tomalin and Woodland, ‘The RT-04 Evaluation Structural Metadata Systems at CUED’, Proc. Fall 2004 RT-04 Workshop.