

Some Results on CTS Quick Transcription and Fisher Data

Phil Woodland and Ricky Chan

5th Sept 2003



Cambridge University Engineering Department

Overview

- Some results on Switchboard 20hr transcriptions
 - different transcriptions
 - ML training
 - discriminative training
- Initial results on 30hr of Fisher data



Sets of Switchboard-1 20hr Data

- MSU transcriptions 20h subset (slow accurate)
 - final processed transcription: 21.7h
 - average sentence length: 3.9s
- LDC 20h quick transcription
 - 5xRT LDC quick transcription
 - as supplied 18.3h, average utt length 2.1s
 - after processing (padding, alignment, joining) 18.3hr/ 4.2s
- BBN/WordWave 20h
 - supplied WordWave transcription: 19.1h, 3.8s
 - after processing: 17.6h, 3.5s



Experimental Results on Quick Transcription: Dev01

- GI HMMs, VTLN, 8 Gaussians/state, 2500 triphone states
- Testsets: dev01, eval03
- Single pass unadapted trigram decoding

	Overall	Swbd1	Swbd2	Cellular	Male	Female
MSU 20h MLE	43.4	33.6	47.9	48.9	43.1	43.7
LDC 20h QT MLE	43.6	33.9	48.3	49.0	43.5	43.7
BBN 20h WWave MLE	43.6	33.7	48.0	49.4	43.7	43.5
MSU 20h MPE	40.5	30.4	45.0	46.3	40.6	40.3
LDC 20h QT MPE	41.2	31.0	45.6	47.1	40.9	41.5
BBN 20h WWave MPE	41.2	31.0	45.9	46.9	41.2	41.2



Experimental Results on Quick Transcription: Eval03

	Overall	Swbd	Fisher	Male	Female
MSU 20h MLE	43.5	49.0	37.6	44.7	42.3
LDC 20h QT MLE	43.8	49.1	38.2	45.3	42.4
BBN 20h WWave MLE	44.0	49.5	38.0	45.5	42.4
MSU 20h MPE	40.5	46.1	34.5	41.9	39.1
LDC 20h QT MPE	41.2	46.7	35.4	42.6	39.8
BBN 20h WWave MPE	41.4	47.4	35.0	42.7	40.1

- Similar increases in error for both WordWave and LDC QT
- Increased error rates on MPE (0.6% to 0.9%) vs MLE (0.2% to 0.5%)



Fisher Data (Initial LDC release)

- Initial release of 30hr of Fisher data fisher data (LDC QT)
- Original transcription - 30.8h; utt length 2.0s
- After processing transcription - 33.1h, utt length 3.2s
- 357 female sides (23.75hrs), only 142 male sides (9.43hrs)
- Same test setup as for 20hr QT expts

Dev01 results:

	Overall	Swbd1	Swbd2	Cellular	Male	Female
LDC 30h Fisher MLE	42.0	34.0	46.7	45.4	42.5	41.4
MSU 20h MLE	43.4	33.6	47.9	48.9	43.1	43.7
LDC 30h Fisher MPE	38.1	30.5	43.0	41.1	38.7	37.5
MSU 20h MPE	40.5	30.4	45.0	46.3	40.6	40.3



- 30hr Fisher (QT) is more effective than 20hr Swbd1 (careful)
- MPE gain is larger than MLE gain
- Gains larger on female test data (biased training)

Eval03 results:

	Overall	Swbd	Fisher	Male	Female
LDC 30h Fisher MLE	40.8	45.7	35.6	42.9	38.7
MSU 20h MLE	43.5	49.0	37.6	44.7	42.3
LDC 30h Fisher MPE	37.0	41.9	31.8	39.1	34.9
MSU 20h MPE	40.5	46.1	34.5	41.9	39.1

- Larger gains on eval03 than on dev01 from Fisher
- Similar relative gains on Swbd2-5 to Fisher ...



Summary

- QT on various sets of Swbd1 transcriptions
 - degradation from MSU 0.2% to 0.5% for MLE training
 - degradation from MSU 0.6% to 0.9% for MPE training
 - error rates from WWave slightly higher than LDC QT for eval03
 - need to make MPE more robust to QT?
- Initial results using Fisher is encouraging
 - Need to scale to using more data rapidly
 - Will be interesting to compare WordWave and LDC QT on Fisher

