Ongoing Experiments with Fisher Data

Ricky Chan, Gunnar Evermann Bin Jia,
David Mrva, Phil Woodland

4th Dec 2003

Cambridge University Engineering Department

Overview

- Initial experiments on using **large** amounts of Fisher data
  - data processing
  - language modelling
  - ML training
  - MPE discriminative training
- Experimental results on
  - h5train03 (360 hours used in CUED 2003 eval system)
  - 500+ hours Fisher
  - combined set
- Evaluated using unadapted and adapted systems

Training and Test Data Sets

- Acoustic training data

  h5train03b 360h data set.
  - 290h LDC data (SwbI, CHE, Swb Cellular) with MSU/LDC careful transcriptions.
  - 70h BBN data (Cellular, Swb2-2) with quick transcriptions

  fisher3896 520h Fisher data set, 3896 conversations

  fisher3896+h5 880h data set, the combined set of h5etrain03b and fisher3896

- Test sets

  eval02 5h set from SwbI, Swb2 and Swb Cellular data, 60 conversations

  eval03 6h set from Fisher and Swb2-5 data, 72 conversations

Fisher data processing

- Original transcriptions: 550h data (424h BBN data, 126h LDC data)
- Normalize the text, joining, padding
- Apply replacement rules
  - Abbreviations, typos, non-speech, ...
  - e.g. FBI → F. B. I., MOULD → MOLD, [NOISE] → -
  - about 2000 replacement rules were produced
- Produce pronunciations for 950 unknown words with frequency greater than 2
- 4900 unknown words remain → remove 10h segments with unknown words
- aligning the segments and fixing silence boundaries
  - 10h segments fail to align
  - 520h fisher data remain (Gender imbalance: 340h female, 180h male)

Acoustic Modelling and Testing

- Acoustic model
  - PLP + VTLN + HLDA front-end
  - cross-word triphone, 6200 tied states
  - 28 variable Gaussian mixture components per state
  - Gender Independent ML and MPE models
- Single Pass unadapted system
  - Trigram LM
  - No adaptation
  - Pruning set for $\sim$ 5xRT
- CU-HTK P1-P2 system (P2 adapted)
  - P1, P2 architecture of CU-HTK 2003 CTS 10xRT eval system
  - Trigram decoding, fourgram lattice rescoring
  - overall $\sim$ 5xRT include adaptation
- Word-list + basic LMs as CTS 2003 eval system

Unadapted single pass decoding WER: Eval03

		Total	Swb2-5	Fisher	Male	Female
h5train03b (360h)	ML	31.9	36.5	27.0	32.8	31.0
ML fisher3896 (520h)	ML	31.2	35.2	26.8	32.8	29.5
ML fisher3896+h5 (880h)	ML	31.0	35.2	26.4	32.4	29.5
MPE h5train03b (360h)	MPE	27.7	32.1	22.9	28.8	26.5
MPE fisher3896 (520h)	MPE	26.4	30.5	22.1	28.3	24.6
MPE fisher3896+h5 (880h)	MPE	25.7	29.9	21.3	27.4	24.1

eval03, trigram LM, unadapted

- fisher3896: performs better than h5train03b, more gain for Swbd2-5 than Fisher, larger gains for Female than Male

- fisher3896+h5: perform better than fisher3896, more gain for Fisher than Swbd, lessens gender imbalance

- Larger gains obtained from MPE than ML with extra data

Unadapted single pass decoding WER: Eval02

	Overall	SwbI	SwbII	SwbC
ML h5train03b (360h)	33.4	27.9	34.6	36.7
ML fisher3896 (520h)	33.4	29.4	34.8	35.5
ML fisher3896+h5 (880h)	32.7	28.3	33.6	35.4
MPE h5train03b (360h)	28.9	24.2	29.6	32.0
MPE fisher3896 (520h)	28.5	25.2	29.5	30.4
MPE fisher3896+h5 (880h)	27.6	23.7	28.1	30.2

eval02, trigram LM, unadapted

- fisher3896: similar overall performance as h5train03b for ML but better for MPE (performs better for SwbC, similar for SwbII, poorer for SwbI)

- fisher3896+h5: performs better than fisher3896, obvious improvements for SwbI and Swb2, minor improvments in SwbC

Revised LM

- LM03: LMs/trainin texts used for 2003 eval
- LM03+Fi3896: LM03 + Fisher3896
- Built separate LMs for each component data source and then interpolate/merge

- Full models also interpolate with 03 eval class-based model (not retrained with Fisher data)

- Interpolation weights for word 4gram LM components for LM03+Fi

BN style	0.18
google	0.08
cell1	0.17
che+swbI	0.20
swbII	0.10
fisher3896	0.26

- Interpolation weights set on dev01, eval00, eval01, eval02 data (no Fisher ...)

Revised LM contd..

- Perplexities on eval02 & eval03 (Swb2 and Fisher subsets)

Language Model	eval03SW	eval03FI	eval03	eval02
full LM03	56.9	59.4	58.1	61.8
full LM03+Fi	55.2	55.7	55.4	60.3
word 4gram LM03+Fi	55.4	55.8	55.6	60.6
fisher3896 only	68.5	65.7	67.2	79.4

- Adding fisher3896 to LM training data decreased the PP of full eval03 LM
  - by 2.7 points (4.6% rel.) on eval03
  - by 3.8 points (6.3% rel.) on Fisher part of eval03

New LMs: Eval03 Unadapted

		Overall	Swbd	Fisher	Male	Female
h5train03b	LM03 tg	27.7	32.1	22.9	28.8	26.5
h5train03b	LM03+Fi	27.2	31.7	22.3	28.2	26.1
fisher3896	LM03	26.4	30.5	22.1	28.3	24.6
fisher3896	LM03+Fi	25.9	30.0	21.5	27.6	24.2
fisher3896+h5	LM03	25.7	29.9	21.3	27.4	24.1
fisher3896+h5	LM03+Fi	25.2	29.5	20.6	26.8	23.5

MPE training, eval03, trigram LM, unadapted

- Consistent 0.5% overall improvement from LM03+Fi

- Both Fisher and Swbd obtain similar improvement from LM03+Fi

New LMs: Eval03 with CU-HTK P1-P2 System

		Overall	Swbd	Fisher	Male	Female
h5train03b	LM03	24.6	28.7	20.2	25.7	23.5
h5train03b	LM03+Fi	23.9	28.2	19.3	25.0	22.8
fisher3896	LM03+Fi	23.1	27.0	18.9	24.6	21.6
fisher3896+h5	LM03+Fi	22.7	26.6	18.5	24.2	21.1

MPE training, eval03, 4-gram LM, adapted

- h5train03b: compare with LM03, LM03+Fi gives 0.7% overall improvement
- fisher3896: performs 0.8% better than h5train03b (LM03+Fi)
- fisher3896+h5: performs 0.4% better than fisher3896 (with LM03+Fi)
- Total 1.9% overall WER reduction adding fisher3896 to h5train03b for both acoustic model and LM training

New LMs: Eval02 with CU-HTK P1-P2 System

		Overall	SwbI	SwbII	SwbC
h5train03b	LM03	26.0	22.0	26.0	29.3
h5train03b	LM03+Fi	25.5	21.8	25.5	28.6
fisher3896	LM03+Fi	25.5	22.7	25.8	27.6
fisher3896+h5	LM03+Fi	25.0	21.6	25.2	27.5

MPE training, eval03, 4-gram LM, adapted

- h5train03b: compare with using LM03, using LM03+Fi gives 0.5% overall improvement

- fisher3896 gives same performance as h5train03b (LM03+Fi)

- fisher3896+h5: performs 0.5% better than fisher3896

- Total 1.0% overall improvement by adding fisher3896 to h5train03b in both acoustic model and language model training data

Summary/Conclusion

- Experiments on 550 hours raw of Fisher data
- Fisher with quick transcription better than 360hour set with mainly careful transcription except for Swb1
- For Fisher subset of eval03
  - For unadapted system, for full training set, MPE training gives more improvement than ML (1.6% vs 0.6%)
  - Adding Fisher training to LM gains 0.6% abs
  - With adaptation, overall 1.9% abs better from adding to LM and acoustic training
- Current results are quick first try: same number of parameters as eval03 training: scope for further improvement