SUMMARY OF PROGRESS AT LIMSI

presented by Jean-Luc Gauvain

EARS RT-04 meeting
Palisades, New York
November 8, 2004

MAJOR TASKS

- CTS English *(J.L. Gauvain, L. Lamel, H. Schwenk, G. Adda, F. Lefevre)*
- BN English *(J.L. Gauvain, L. Lamel, H. Schwenk, G. Adda, F. Lefevre)*
- BN Non English *(J.L. Gauvain, L. Lamel, A. Messaoudi, L. Chen)*
- Speaker diarization *(C. Barras, S. Meignier, X.Zhu, J.L. Gauvain)*

GENERAL IMPROVEMENTS

- Better acoustic modeling (speaker adaptive training, full-covariance for state-tying, semi-tied covariance models)
- Faster software and infrastructure for acoustic model training on large corpora (light supervision for BN)
- Improved acoustic model adaptation
- Updated language models and dictionaries for all tasks
- Speed up training and decoding for neural network language models
- Faster decoding for CTS and BN
- Large improvement on speaker diarization task

PROGRESS ON CTS ENGLISH

- Better acoustic models (1.7% absolute reduction)
- Improved acoustic and language models and lexicon using Fisher training data (2.5% absolute)
- Faster and better decoding with AM adaptation (0.4% absolute with a factor of 6 speed-up)
- Multiple phone set modeling (0.4 - 0.7% absolute)
- Cascade and Rover combination with BBN
- Overall relative error reduction without Rover combination of about 23%

PROGRESS ON ENGLISH BN

- Focused on cross-site combination
- Integrated feature optimization (MLLT) for diagonal Gaussians and speaker adaptive training (SAT)
- Improved unsupervised adaptation (CMLLR, MLLR, with variable number of adaptation classes)
- Training on 600 hours of BN data, with light supervision
- Neural network continuous space LM
- About 20% relative WER reduction since RT03

NEURAL NETWORK LM

- Training speed-up (better algorithms and code tuning)

  Factor of 30 speed-up

- Recognition speed-up (improved lattice rescoring)

  0.6xRT $\longrightarrow$ 0.06xRT

  including lattice expansion and consensus decoding

- Word error reduction relative to the 4-gram backoff LM

  CTS dev04: 16.0% $\longrightarrow$ 15.5% (B1-L1, 6xRT)

  BN dev04: 10.4% $\longrightarrow$ 10.1% (B1-L1, 6xRT)

  Trained on about 27M words for BN and CTS

SYSTEM COMBINATION

- Two decoding strategies for cross-site system combination
- Lattice rescoring:
  - Faster (allows computation to be shared between systems)
  - Need to deal with compound words and OOVs
  - Reduces combination gain (due to sharing of components)
- Full search:
  - Requires fast decoding
  - Results in higher WER, but comes out ahead after combination
- Cross-site adaptation with fast full decode found to be more efficient for the BBN/LIMSI and SuperTeam combinations

PROGRESS ON NON-ENGLISH BN

Mandarin

- Improved acoustic models with lightly supervised training with TDT
  Word error reduction of over 10% on dev03 and eval03

Arabic

- Explicit modeling of short vowels (AMs, lexicon)
- 65k word class 4-gram language model regrouping all vowelized forms for each non-vowelized entry
- Dictionary has over 400k vowelized forms, obtained semi-automatically
- Acoustic model training on 150 hours of data
- RT-04F Primary system WER 18.5%, Contrast system 20.2%

PROGRAM-WIDE CONTRIBUTIONS

- LIMSI delivered aligned ASR transcripts with CC to LDC for fast manual correction
- Led design of the STT BN dev'04 set (targeting progress set WER)
- Collaborated with CU to produced MDE alignments for Eval03S and Eval03F data
- Alignment software packaged for NIST
  h4align, h5align, 150k dictionary
- STT website (slides of STT meetings, information exchange)
- Participation in weekly STT teleconferences

MAJOR ACHIEVEMENTS

- Participated in 4 RT-04F tasks (5 systems):
  CTS and BN English, BN Arabic, Speaker Diarization
- CTS English
  - Reduced WER by 23% and factor of 6 speed-up
  - Efficient BBN+LIMSI cascade/Rover system combination
  - BBN+LIMSI system 13.5% on progress test
- BN English
  - Concentrated effort on system combination
  - Reduced relative WER by about 20%
  - Efficient BBN/LIMSI combination: 9.5% on progress test
  - Also part of SuperTeam: 8.6% on progress test
- LIMSI STT partitioner output used by 4 of the 6 primary BN English RT-04 submissions

RT-04 TECHNICAL TALKS

- The RT04 BBN/LIMSI 20xRT English CTS System *(J.L. Gauvain, R. Prasad)*
- The RT04 BBN/LIMSI 10xRT BN English System *(L. Nyugen, L. Lamel)*
- Alternate phone models for CTS *(L.Lamel)*
- Using neural network language models for LVCSR *(H. Schwenk)*
- Improving Speaker Diarization *(C. Barras)*
- Towards using STT for Broadcast News Speaker Diarization *(L.Lamel)*
- The LIMSI RT04 BN Arabic System *(J.L. Gauvain)*

CONCLUSIONS

- Good progress on all tasks
- Strong teamwork with BBN (also SuperTeam participation)
- Participated in 4 RT-04F tasks (5 systems)
- Met program targets
- Still porting some CTS improvements to BN
- Training on large corpora: major effort, still in progress
- Some research did not make the RT-04 eval (not ready for integration, CPU time constraints)
  examples: dynamic LM, duration model, full covariance, syllable-position based phone modeling, discriminative LM training