A General Artificial Neural Network Extension for HTK

Chao Zhang & Phil Woodland

University of Cambridge

Natural Speech Technology

Edinburgh – Cambridge – Sheffield

15 April 2015

Overview

- Design Principles
- Implementation Details
  - Generic ANN Support
  - ANN Training
  - Data Cache
  - Other Features
- A Summary of HTK-ANN
- HTK based Hybrid/Tandem Systems & Experiments
  - Hybrid SI System
  - Tandem SAT System
  - Demo Hybrid System with Flexible Structures
- Conclusions

Design Principles

- The design should be as generic as possible.
    - Flexible input feature configurations.
    - Flexible ANN model architectures.
- HTK-ANN should be compatible with existing functions.
    - To minimise the effort to reuse previous source code and tools.
    - To simplify the transfer of many technologies.
- HTK-ANN should be kept “research friendly”.

Generic ANN Support

- In HTK-ANN, ANNs have layered structures.
  - An HMM set can have any number of ANNs.
  - Each ANN can have any number of layers.
- An ANN layer has
  - Parameters: weights, biases, activation function parameters
  - An input vector: defined by a **feature mixture** structure
- A feature mixture has any number of **feature element**s
- A feature element defines a fragment of the input vector by
  - Source: acoustic features, augmented features, output of some layer.
  - A context shift set: integers indicated the time difference.

Generic ANN Support

- In HTK-ANN, ANN structures can be any directed cyclic graph.
- Since only standard EBP is included at present, HTK-ANN can train non-recurrent ANNs properly (directed acyclic graph).

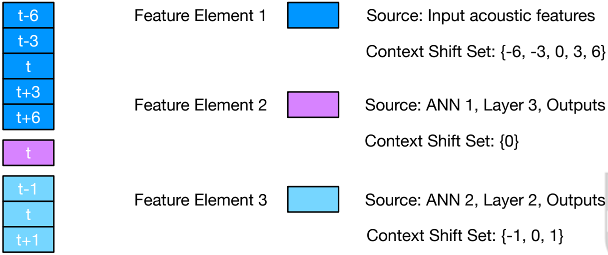

Figure: An example of a feature mixture.

UNIVERSITY OF CAMBRIDGE

ANN Training

- HTK-ANN supports different training criteria
  - Frame-level: CE, MMSE
  - Sequence-level: MMI, MPE, MWE
- ANN model training labels can come from
  - Frame-to-label alignment: for CE and MMSE criteria
  - Feature files: for autoencoders
  - Lattice files: for MMI, MPE, and MWE criteria
- Gradients for SGD can be modified with momentum, gradient clipping, weight decay, and max norm.
- Supported learning rate schedulers include List, Exponential Decay, AdaGrad, and a modified NewBob.

Data Cache

Natural Speech Technology

- HTK-ANN has three types of data shuffling
  - Frame based shuffling: CE/MMSE for DNN, (unfolded) RNN
  - Utterance based shuffling: MMI, MPE, and MWE training
  - Batch of utterance level shuffling: RNN, ASGD

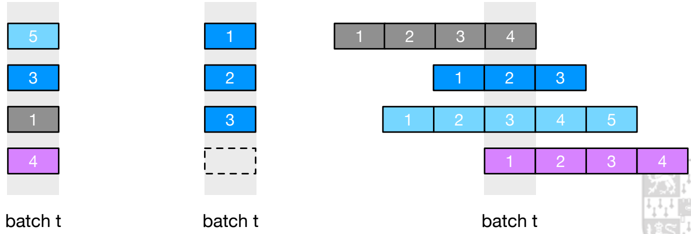

Figure: Examples of different types of data shuffling.

Other Features

Natural Speech Technology

- Math Kernels: CPU, MKL, and CUDA based new kernels for ANNs
- Input Transforms: compatible with HTK SI/SD input transforms
- Speaker Adaptation: an ANN parameter unit online replacement
- Model Edit
  - Insert/Remove/Initialise an ANN layer
  - Add/Delete a feature element to a feature mixture
  - Associate an ANN model to HMMs
- Decoders
  - HVite: tandem/hybrid system decoding/alignment/model marking
  - HDecode: tandem/hybrid system LVCSR decoding
  - HDecode.mod: tandem/hybrid system model marking
  - A Joint decoder: log-linear combination of systems (same decision tree)

UNIVERSITY OF CAMBRIDGE

A Summary of HTK-ANN

- Extended modules: HFBLat, HMath, HModel, HParm, HRec, HLVRec
- New modules
  - HANNet: ANN structures & core algorithms
  - HCUDA: CUDA based math kernel functions
  - HNCache: Data cache for data random access
- Extended tools: HDecode, HDecode.mod, HHEd, HVite
- New tools
  - HNForward: ANN evaluation & output generation
  - HNTrainSGD: SGD based ANN training

Building Hybrid SI Systems

- Steps of building CE based SI CD-DNN-HMMs using HTK
  - Produce desired tied state GMM-HMMs by decision tree tying (HHEd)
  - Generate ANN-HMMs by replacing GMMs with an ANN (HHEd)
  - Generate frame-to-state labels with a pre-trained system (HVite)
  - Train ANN-HMMs based on CE (HNTrainSGD)
- Steps for CD-DNN-HMM MPE training
  - Generate num./den. lattices (HLRescore & HDecode)
  - Phone mark num./den. lattices (HVite or HDecode.mod)
  - Perform MPE training (HNTrainSGD)

ANN Front-ends for GMM-HMMs

Natural Speech Technology

- ANNs can be used as GMM-HMM front-ends by using a feature mixture to define the composition of the GMM-HMM input vector.
- HTK can accomodate a tandem SAT system as a single system
  - Mean and variance normalisations are treated as activation functions.
  - SD parameters are replaceable according to speaker ids.

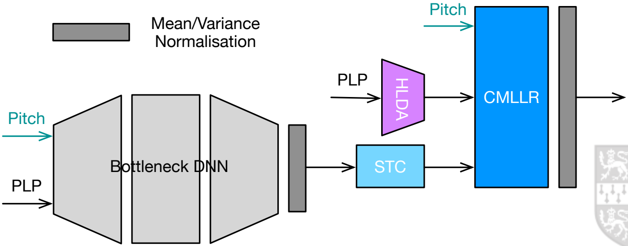

Figure: A composite ANN as a Tandem SAT system front-end.

UNIVERSITY OF CAMBRIDGE

Standard BOLT System Results

- Hybrid DNN structure: $504 \times 2000^4 \times 1000 \times 12000$
- Tandem DNN structure: $504 \times 2000^4 \times 1000 \times 26 \times 12000$

System	Criterion	%WER
Hybrid SI	CE	34.5
Hybrid SI	MPE	31.6
Tandem SAT	MPE	33.2
Hybrid SI $\otimes$ Tandem SAT	MPE	31.0

Table: *Performance of BOLT tandem and hybrid systems with standard configurations evaluated on dev'14. $\otimes$ is the joint decoding with system dependent combination weights (1.0, 0.2).*

WSJ Demo Systems with Flexible Structures

Natural Speech Technology

- Stacking MLPs: $(468 + (n-1) \times 200) \times 1000 \times 200 \times 3000$, $n = 1, 2, \ldots$. Each MLP takes all previous BN features as input.
- The top MLP does not have a BN layer.
- System was trained with CE based discriminative pre-training and fine-tuning.
- Systems were trained with 15 hours Wall Street Journal (WSJ0).

FNN Num	%Accuracy		%WER	
	Train	Held-out	65k dt	65k et
1	69.9	58.1	9.3	10.9
2	72.8	59.1	9.0	10.4
3	73.9	59.1	8.8	10.7

Table: *Performance of the WSJ0 Demo Systems.*

UNIVERSITY OF CAMBRIDGE

Conclusions

- HTK-ANN integrates native support of ANNs into HTK.
- HTK based GMM technologies can be directly applied to ANN-based systems.
- HTK-ANN can train FNNs with very flexible configurations
  - Topologies equivalent to DAG
  - Different activation functions
  - Various input features
  - Frame-level and sequence-level training criteria
- Experiments on 300h CTS task showed HTK can generate standard state-of-the-art tandem and hybrid systems.
- WSJ0 experiments showed HTK can build systems with flexible structures.
- HTK-ANN will be available with the release of HTK 3.5 in 2015.