UNBIASED SEMI-SUPERVISED LF-MMI TRAINING USING DROPOUT

SIBO TONG, APOORV VYAS, PHILIP N. GARNER, HERVÉ BOURLARD
IDIAP RESEARCH INSTITUTE, MARTIGNY, SWITZERLAND
ECOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE (EPFL), SWITZERLAND
{sibo.tong, apoorv.vyas, phil.garner, bourlard}@idiap.ch

OVERVIEW

Goals

- Unbiased Semi-supervised training for LF-MMI

Our approach

- Train a seed model using supervised data.
- Use seed model to generate supervision lattice for untranscribed data
  - Sample N hypotheses with dropout on and combine
- Minimize the MMI Loss on generated lattices

Main Result

- Fisher English: WER recovery of ~ 51.6% over regular semi-supervised LF-MMI training

SEMI-SUPERVISED TRAINING: FLOW CHART

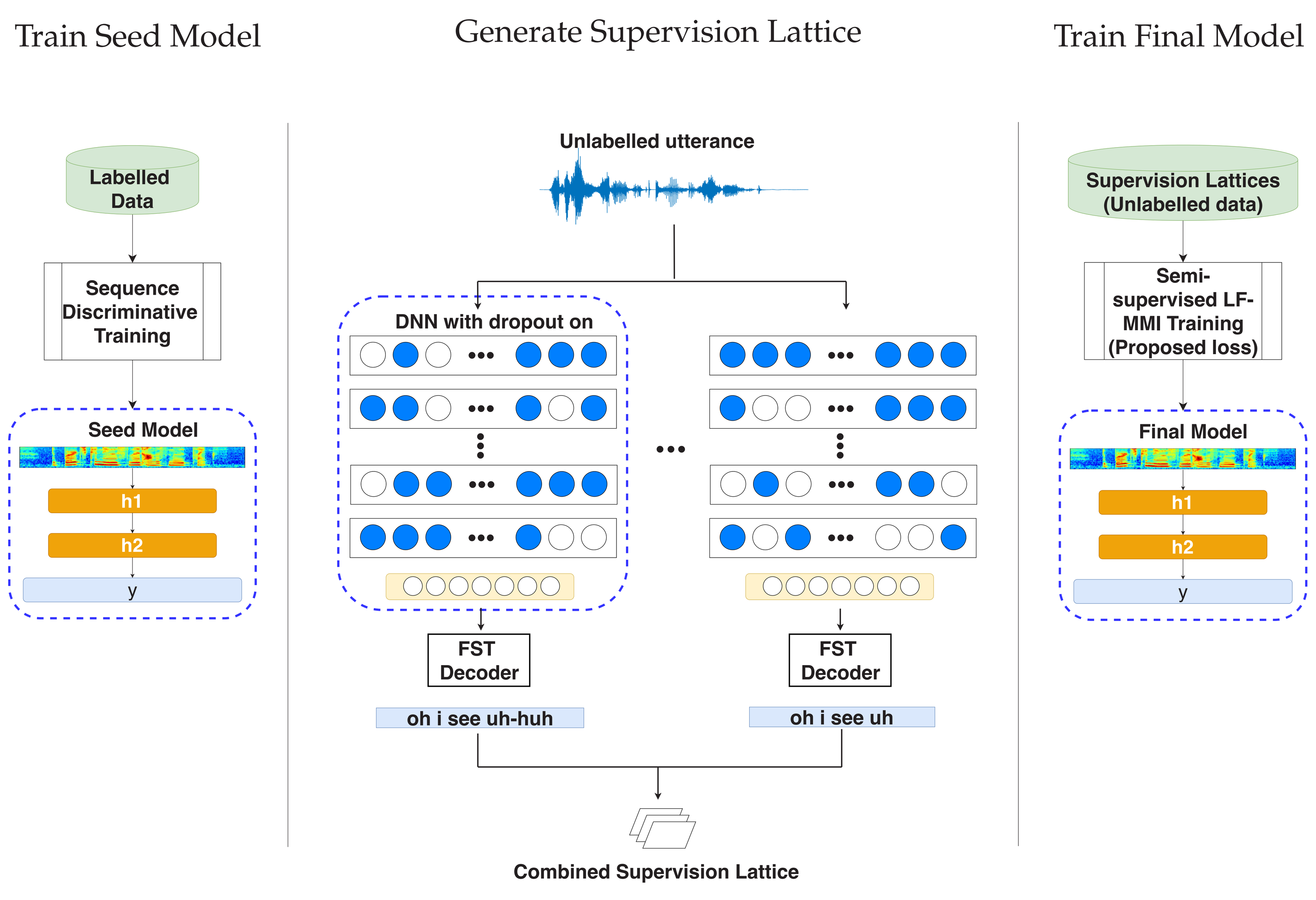

Figure 1: Flow-chart of the proposed method. Each network in the step 2 represents a random selection of the nodes. The white nodes denote dropped out units.

SEMI-SUPERVISED TRAINING: LOSS FUNCTION

Regular semi-supervised loss:

$$\mathcal{L}_{\text{MMI}} = \max_{\theta} \sum_{u=1}^{U} \log \left(\sum_{W \in \mathcal{G}^{(u)}_{\text{num}}} P(W|\mathbf{O}^{(u)}, \theta) \right)$$

Our proposed loss:

$$\mathcal{L}_{\text{p}} = \max_{\theta} \sum_{u=1}^{U} \log \left(\mathop{\mathbf{E}}_{\mathbf{W} \sim P(W|O^{u}, \theta_s)} P(W|\mathbf{O}^{(u)}, \theta) \right)$$

- The regular semi-supervised LF-MMI loss can be seen as an approximation to the proposed loss with the equally likely word-sequences sampled from the decoding lattice

EXPERIMENT SET UP

Dataset: Fisher English corpus

- Supervised (50 hours)
- Unsupervised (250 hours)

Acoustic Models:

- TDNN: 8 hidden layers, 450 hidden units, 0.2 dropout

Metric:

- Word Recovery Rate (WRR) = $\dfrac{\text{BaselineWER} - \text{SemisupWER}}{\text{BaselineWER} - \text{OracleWER}}$

Baselines:

- Lattice based semi-supervised training

RESULT: DROPOUT-OFF VS DROPOUT-ON (QUALITATIVE)

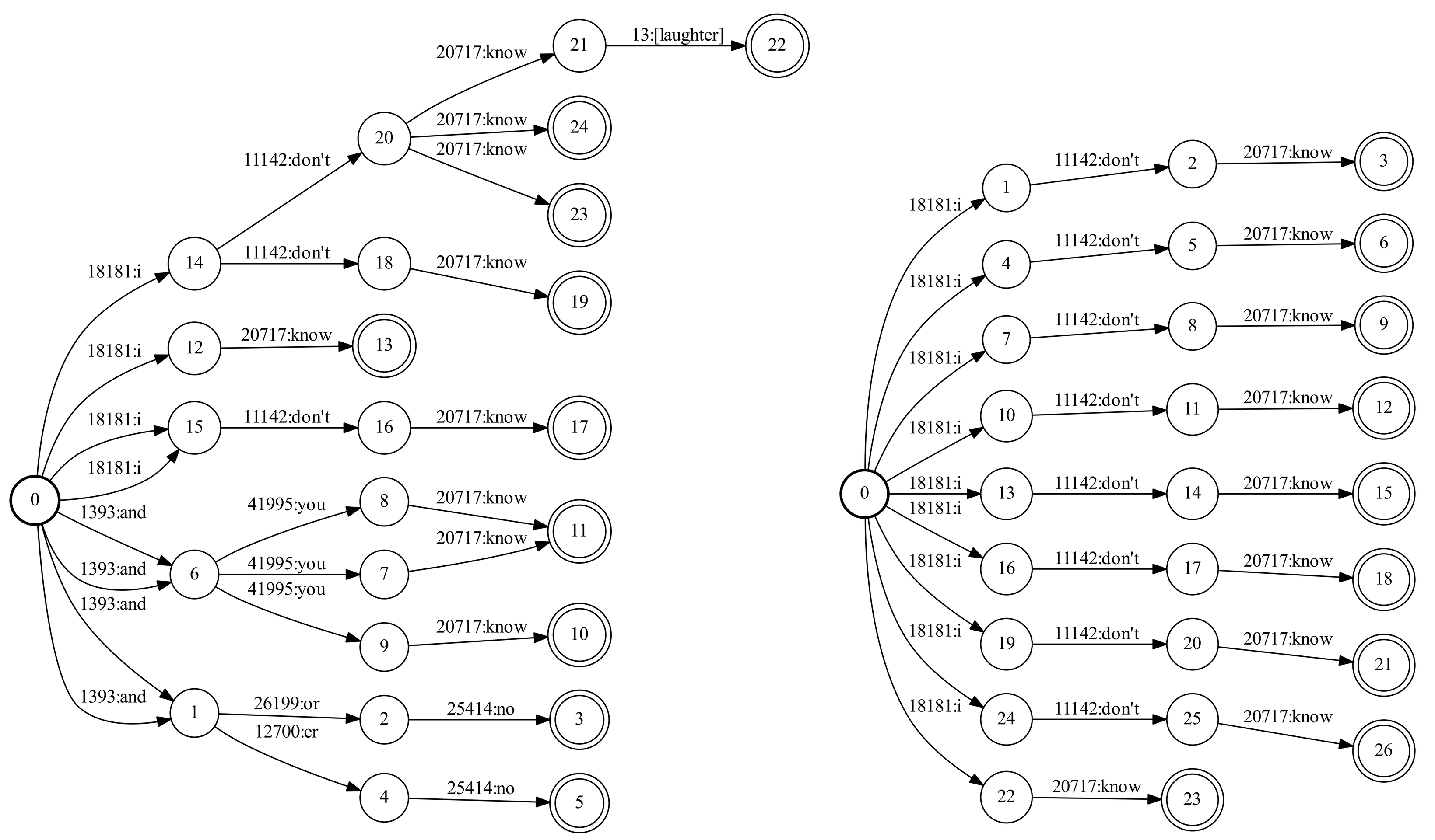

Figure 2: Lattices of a clearly spoken utterance. (a) pruned decoding lattice from a dropout-off acoustic model. (b) unbiased lattice from multiple dropout decoding samples.

- Decoding lattice with dropout-off can bias the training towards incorrect paths which deteriorates the supervision

RESULT: DROPOUT-OFF VS DROPOUT-ON (QUANTITATIVE)

	avg. WER	SER
Regular Lat	23.6	87.8
Lat-comb	23.1	75.7

Table 1: *Comparing averaged Word Error Rate (WER %) and Sentence Error Rate (SER %) between combined and regular decoding lattice.*

- Better WER and SER indicate that dropout lattice helps reduce the effect of incorrect hypotheses when the model is confident
- Alternative paths are explored when the acoustic model is uncertain

RESULTS: ANALYSIS (NUMBER OF DROPOUT SAMPLES N)

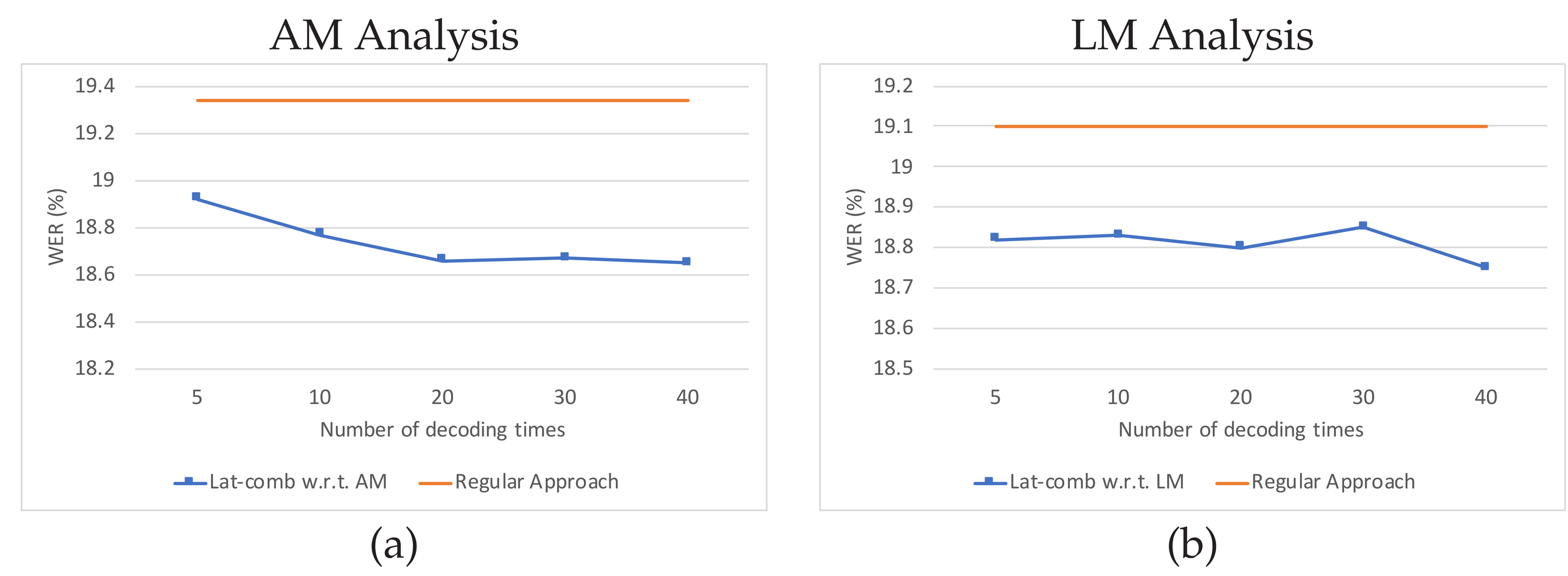

Figure 3: WER (%) of different semi-supervised training setup by varying the value of dropout samples N for (a) Acoustic Model (AM) only (b) Language Model (LM) only. The red line denotes the regular semi-supervised training approach.

- AM dropout analysis: Fixed N-Gram LM is used
- LM dropout analysis: RNN based LM with dropout on. AM dropout is off

RESULTS (WER ESTIMATION)

System	Dev	Test	WRR
50h supervised	21.0	20.9	-
Regular Approach	19.1	19.2	53.7 %
Lat-comb w.r.t. AM	18.5	18.3	76.1%
Lat-comb w.r.t. LM	18.8	18.7	65.7%
Lat-comb w.r.t. AM+LM	**18.5**	**18.2**	77.6%
Oracle	17.7	17.5	

Table 2: Comparison between combined lattice and regular decoding lattice in WER(%). The 50h supervised system is used as baseline to calculate WRR.

- Most gains come from acoustic dropout
- Language model dropout provides mild improvement

CONCLUSIONS & FUTURE WORK

- Semisupervised LF-MMI training with dropout on can be seen as unbiased risk minimization under uncertainty
- Dropout sampling can be applied to both AM and LM to improve the WER over the regular semi-supervised training framework
- In the future, we intend to extend to idea to other frameworks such as end-to-end training, and also work on reducing the time required for multiple times of decoding

ACKNOWLEDGMENT

This research was supported by Swiss National Science Foundation project SHISSM, grant agreement 200021-175589, and the European Community H2020 SUMMA project No. 688139