Dropout Approaches for LSTM Based Speech Recognition Systems

Jayadev Billa | jbilla@isi.edu
Information Sciences Institute, USC Viterbi School of Engineering

USC Viterbi School of Engineering

CONTRIBUTION

- Empirical study on dropout in LSTM-CTC ASR systems
- Introduces approaches to combine different dropout paradigms
- Virtually eliminates performance gap between LSTM-CTC & hybrid DNN systems with limited data

SYSTEM	%WER
Librispeech 100hr best (dev-clean, this paper)	7.37
Kaldi (p-norm DNN, LDA+MLLT+SAT, 100hrs)	7.91
Kaldi (p-norm DNN, LDA+MLLT+SAT, 460hrs)	7.16
GALE Arabic 250hr best (this paper)	19.70
Kaldi (TDNN chain model, MMI, 415hrs)	20.26
Kaldi (TDNN/LSTM chain model, MMI, 415hrs)	17.64

PRIOR WORK

- Until recently, limited success with dropout in ASR
- Many options proposed
  - Apply on feed forward connections (early attempts…)
  - Apply within recursion — but where?

We systematically explore and extend this earlier work

BASE SYSTEM

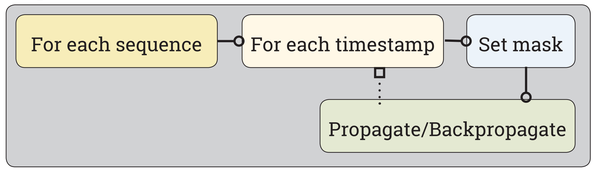
- EESEN based LSTM-CTC ASR
  - 40 mel-fbank features + Δ + ΔΔ, trigram LM
  - Max perturbation data augmentation

SYSTEM	%WER
Librispeech 100hr (dev-clean)	9.78
GALE Arabic 250hr (withheld test)	22.84

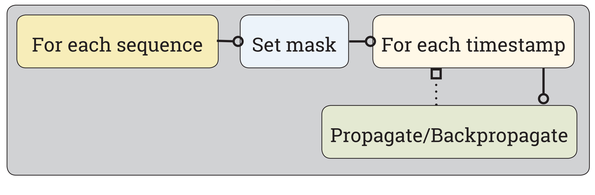
- Dropout enabled systems use stacked and strided (SS) features

SEQUENCE VS TIME-STEP

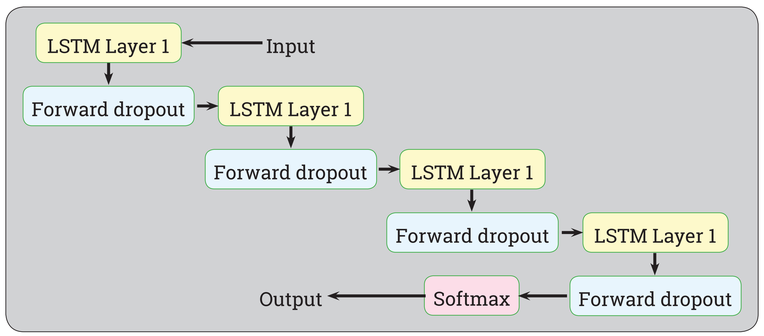
- Time step



- Alternative: fix mask across sequence



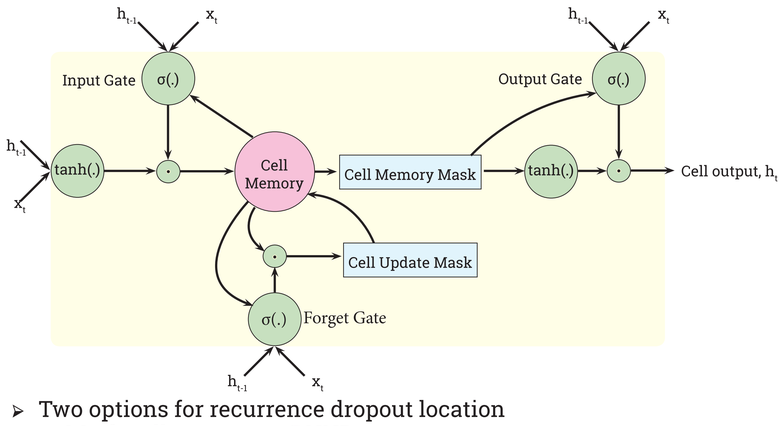
FEEDFORWARD DROPOUT



- **%WER:** Baseline (SS) > Baseline > Forward step > Forward sequence

LIBRISPEECH 100HR SYSTEM	%WER	%RI
Baseline	9.78	—
Baseline (SS)	10.72	-9.61
Forward step	9.17	6.24
Forward sequence	8.63	11.76

RECURRENT DROPOUT



- Two options for recurrence dropout location
  - Mask cell memory — RNNDrop
  - Mask cell update — Recurrent dropout w/o memory loss (NML)

LIBRISPEECH 100HR SYSTEM	%WER	%RI
RNNDrop step	9.07	7.26
NML step	8.55	12.58
NML sequence	8.45	13.60

- **%WER:** Baseline (SS) > Baseline > NML Step ≈ NML Sequence

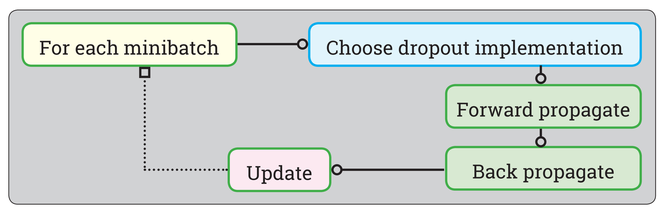
Dropout code and Librispeech experiments available in Eesen Github repo @ https://github.com/srvk/eesen

COMBINING DROPOUT

- **Naive** — apply forward and recurrent dropout all the time

LIBRISPEECH 100HR SYSTEM	%WER	%RI
RNNDrop step + Forward step	8.60	12.07
RNNDrop step + Forward sequence	8.85	9.51
NML step + Forward step	8.08	17.38
NML step + Forward sequence	**7.76**	**20.65**
NML sequence + Forward step	**7.72**	**21.06**
NML sequence + Forward sequence	7.97	18.51

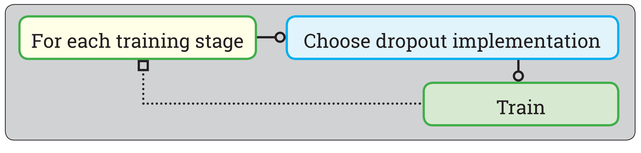
- **Stochastic** — decide which dropout to apply per minibatch



LIBRISPEECH 100HR SYSTEM	%WER	%RI
NML step + Forward step	8.76	10.43
NML step + Forward sequence	8.02	18.00
NML sequence + Forward step	7.86	19.63
NML sequence + Forward sequence	**7.44**	**23.93**

GALE ARABIC 250HR SYSTEM	%WER	%RI
Baseline	22.84	—
NML sequence + Forward sequence	19.70	13.75

- **Cascade** — Train in stages with different dropout type in each stage



LIBRISPEECH 100HR SYSTEM	%WER	%RI
NML sequence + Forward step ①	7.72	21.06
NML sequence + Forward sequence ②	7.97	18.51
Cascade ①➡②	**7.37**	**24.64**

- A dropout schedule is a particular instance of cascade combination
- Stochastic and cascade combination are complimentary.