Bi-directional Lattice Recurrent Neural Networks for Confidence Estimation

Qiujia Li*, **Preben Ness***, **Anton Ragni**, **Mark Gales**

Machine Intelligence Laboratory, Cambridge University Engineering Department

Confidence Scores

Introduction

Confidence scores are a measure of uncertainty regarding ASR system output at phone / word / utterance level

Evaluation Metrics

Normalised Cross Entropy (NCE)

– for applications where the *absolute value* matters

$$\text{NCE}(\mathbf{c}, \mathbf{c}^*) = \frac{H(P_c \cdot \mathbf{1}, \mathbf{c}^*) - H(\mathbf{c}, \mathbf{c}^*)}{H(P_c \cdot \mathbf{1}, \mathbf{c}^*)}$$

- P_c is the empirical estimate of ASR correctness
- $\mathbf{c}/\mathbf{c}^*$ is the estimated/reference confidence scores
- $H(\cdot, \cdot)$ is cross-entropy between two sequences

Area Under Precision-Recall Curve (AUC)

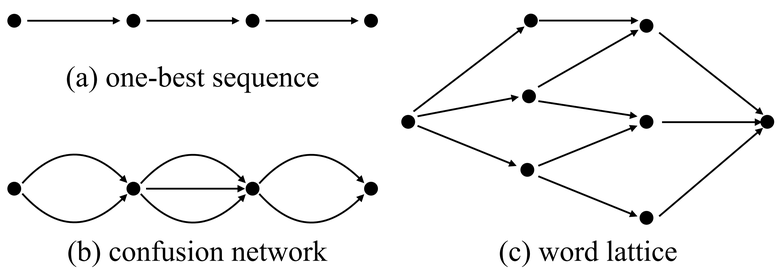
– for applications where the *rank ordering* matters

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

TP = true positives FP = false positives FN = false negatives

ASR Structured Data

- ASR hypothesis can be structured as sequences, confusion networks (CNs) or lattices (Figure 1)
- Word and decoding information are associated with each arc



(a) one-best sequence (b) confusion network (c) word lattice

Figure 1: Standard ASR outputs

Arc Tagging

- Approximate scheme using time overlap (Figure 2)
- Exact scheme using Levenshtein edit distance

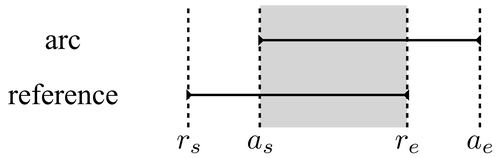
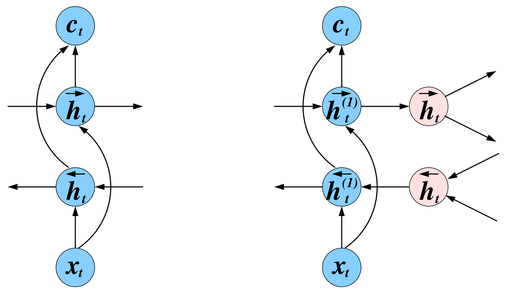


Figure 2: Approximate arc tagging scheme

Bi-directional Lattice Recurrent Neural Networks (BiLatRNNs)

- Bi-directional RNNs (BiRNNs) operate on a sequence of input features $\mathbf{x}_{1:N}$ and model past and future information
- To accommodate structured data, *i.e.* CNs and lattices, BiLatRNNs are introduced (Figure 3)



(a) BiRNN (b) BiLatRNN

Figure 3: Bi-directional neural networks for confidence estimation

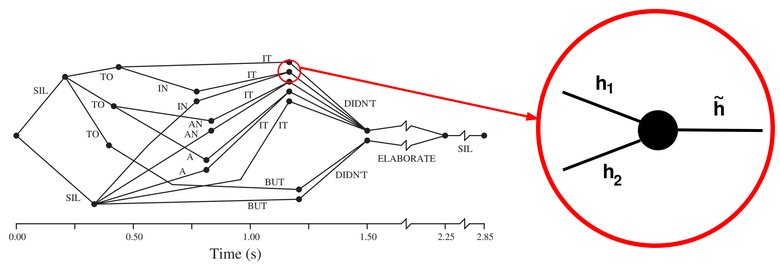


Figure 4: Merging of hidden state vectors in a lattice

Forward-backward algorithm

$$\overrightarrow{h}^{(i)}_{t+1} = \overrightarrow{h}_t x^{(i)}_{t+1}, \text{ where } \overrightarrow{h}_t = \sum_j \alpha_{i,j} \overrightarrow{h}^{(j)}_t$$

Here $\alpha_{i,j}$ is the fixed transition probability from state i to j

Extended to BiLatRNN

$$\overrightarrow{\mathbf{h}}_t = \sum_i \alpha^{(i)}_t \overrightarrow{\mathbf{h}}^{(i)}_t, \qquad \overleftarrow{\mathbf{h}}_t = \sum_i \beta^{(i)}_t \overleftarrow{\mathbf{h}}^{(i)}_t$$

where α, β are the weights associated with each incoming arc.

- Alternatives for merging arcs include a **posterior weighted average** or an **attention mechanism**

Experimental Setup

- Model: 1 layer LSTM + 1 layer feed forward
- Input feature: 50-D word embedding, mapped posterior, duration, acoustic and language model scores

Experimental Results

BiRNN on 1-best

Estimator	NCE	AUC
1-best CN posteriors	-0.1978	0.9081
+decision tree	0.2755	0.9081
+BiRNN	**0.2947**	**0.9197**

Table 1: Confidence estimation on 1-best CN arcs

Confusion Networks

Estimator	NCE	AUC
all CN posteriors	0.3105	0.8243
+decision tree	0.4659	0.8243
+BiLatRNN	**0.4970**	**0.8365**

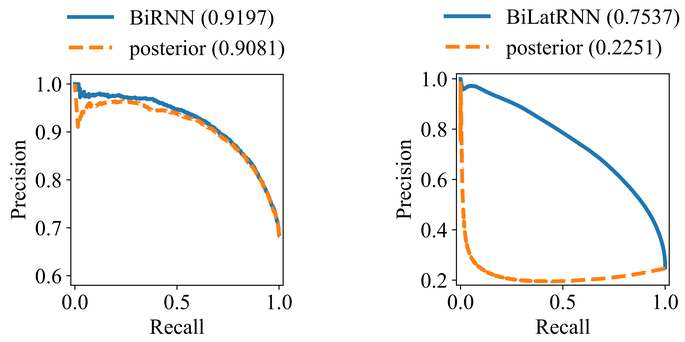
Table 2: Confidence estimation on all CN arcs

Lattices

Estimator	Merge	NCE	AUC
all lattice arc posteriors	–	-5.0386	0.2251
+decision tree	–	-0.0889	0.2251
+BiLatRNN	max	0.3774	0.7453
	mean	0.3788	0.7424
	posterior	0.3880	0.7507
	attention	**0.3921**	**0.7537**

Table 3: Confidence estimation on all lattice arcs

Improvement on AUC performance



(a) 1-best CN arcs (b) all lattice arcs

Figure 5: Bi-directional neural networks for confidence estimation

Conclusions

- BiLatRNNs show significant gains in confidence estimation over all arcs in CNs and lattices, evaluated under either metric
- Many applications will benefit from the improved confidence measure, *e.g.* information retrieval, keyword spotting, speaker adaptation, semi-supervised training

E-Mail: {ql264, pmn26, ar527, mjfg}@eng.cam.ac.uk
Code https://github.com/qiujiali/lattice_rnn