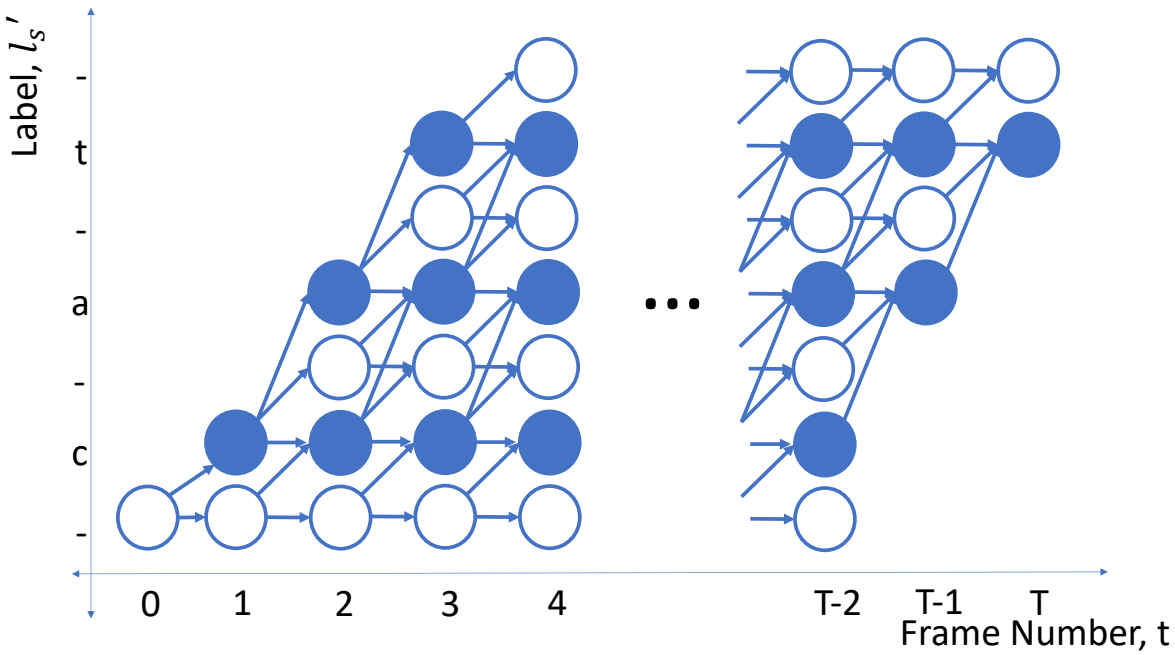
Grad Seminar Week 3: Overview of Paper 1 Example Paper: Connectionist Temporal Classification

Please fill in your name here!

University of Illinois

Please put the date of your presentation here!

CTC Forward Algorithm: Schematic



In order for the neural net to generate a particular label at a particular frame,

- If label l_s is a blank, then the preceding frame must be blank or non-blank.
- If label l_s is non-blank, then it could be preceded by (1) a repetition of itself, or (2) a preceding blank, or (2) the preceding non-blank.

Algorithm Equation, and Definition of its Variables

The neural net is trained to maximize $p(\mathbf{l}|\mathbf{x})$, which is computed as

$$p(\mathbf{l}|\mathbf{x}) = \alpha_T(|\mathbf{l}'|) + \alpha_T(|\mathbf{l}'| - 1)$$

where

$$\alpha_t(s) = \begin{cases} (\alpha_{t-1}(s) + \alpha_{t-1}(s-1))\, y^t_{\mathbf{l}'_s} & \mathbf{l}'_{s-2} = \mathbf{l}'_s \\ (\alpha_{t-1}(s) + \alpha_{t-1}(s-1) + \alpha_{t-1}(s-2))\, y^t_{\mathbf{l}'_s} & \text{otherwise} \end{cases}$$

- $\mathbf{l}$ = output label sequence, $\mathbf{x}$ =input audio frame sequence, $\mathbf{l}'$ = output labels with one blank inserted between each pair of characters, e.g., "-,c,-,a,-,t,-".
- $\alpha_t(s) = p(\mathbf{l}'_{1:s}|\mathbf{x}_{1:t})$ = probability of the first s labels, $\mathbf{l}'_{1:s}$, being the correct output for the first t frames, $\mathbf{x}_{1:t}$.
- $y^t_{\mathbf{l}'_s} = p(\mathbf{l}'_s|\mathbf{x}_t)$ = probability the neural net computes locally for label $\mathbf{l}'_s$ in frame t.

Algorithm Equation: One step of the derivation

First, start by defining more precisely what it means that $\alpha_t(s) =$ probability of the first s labels occurring during the first t frames. Let's define

$$\alpha_t(s) = P(\mathbf{l}'_{1:s}|\mathbf{x}_{1:t})$$

where $\mathbf{l}'_{1:s}$ is the labels up through the sth label, and $\mathbf{x}_{1:t}$ is audio up through the tth frame. Then

$$\begin{aligned}\alpha_t(s) &= \left(P(\mathbf{l}'_{1:(s-1)}|\mathbf{x}_{1:(t-1)}) + P(\mathbf{l}'_{1:s}|\mathbf{x}_{1:(t-1)})\right) P(\mathbf{l}_s|\mathbf{x}_t) \\ &= (\alpha_{t-1}(s-i) + \alpha_{t-1}(s))\, y^t_{\mathbf{l}'_s}.\end{aligned}$$

The two terms inside the parentheses are because the previous label could be different $(s-1)$, or the current label might be a repetition of the previous label (s).