TTIC 31230, Fundamentals of Deep Learning

David McAllester, Winter 2019

Connectionist Temporal Classification (CTC)

Connectionist Temporal Classification (CTC)

A Successful Deep Latent Variable Model

A speech signal

$$x = x_1,\ \ldots,\ x_T$$

is labeled with a phone sequence

$$y = y_1, \ldots, y_N$$

with $N << T$ and with $y_n \in \mathcal{P}$ for a set of phonemes $\mathcal{P}$.

The length N of y is not determined by x and the alignment between x and y is not given.

The CTC Model

$$P_\Phi(y|x) = \sum_z P_\Phi(z|x) P_\Phi(y|z).$$

Input Signal:	$x = x_1, \ldots, x_T$
Latent Label:	$z = z_1, \ldots, z_T, \quad z_t \in \mathcal{P} \cup \{\perp\}$
Output:	$y(z) = y_1, \ldots, y_N \qquad N << T$

$y(z)$ is the result of removing all the occurrences of $\perp$ from z:

$$z \qquad\qquad \Rightarrow \qquad y$$

$$\perp, a_1, \perp, \perp, \perp, a_2, \perp, \perp, a_3, \perp \quad \Rightarrow \quad a_1, a_2, a_3$$

The CTC Model

For $z \in \mathcal{P} \cup \{\perp\}$ we have an embedding $e(z)$. The embedding is a parameter of the model.

$$h_1, \ \ldots, \ h_T = \mathrm{RNN}_\Phi(x_1, \ \ldots, \ x_T)$$

$$P_\Phi(z_t | x_1, \ldots, x_T) = \operatorname*{softmax}_z \ e(z)^\top h_t$$

$z_1, \ \ldots \ z_T$ are all independent given x.

Dynamic Programming

$$\begin{aligned}
x &= x_1, \ \ldots, \ x_T \\
z &= z_1, \ \ldots, \ z_T, \quad z_t \in \mathcal{P} \cup \{\perp\} \\
y &= y_1, \ \ldots, \ y_N, \quad y_n \in \mathcal{P}, \quad N << T \\
y(z) &= (z_1, \ \ldots, \ z_T) - \perp
\end{aligned}$$

$$\begin{aligned}
\vec{y}_t &= (z_1, \ \ldots, \ z_t) - \perp \\
F[n, t] &= P(\vec{y}_t = y_1, \ \ldots, \ y_n) \\
P(y) &= F[N, T]
\end{aligned}$$

Dynamic Programming

$$\vec{y}_t = (z_1, \ \ldots, \ z_t) - \perp$$
$$F[n,t] = P(\vec{y}_t = y_1, \ \ldots, \ y_n)$$

$F[0,0] = 1$

For $n = 1, \ldots, N$ $F[n,0] = 0$

For $t = 1, \ldots, T$

$\quad F[0,t] = P(z_t = \perp)F[0,t-1]$

$\quad$ for $n = 1, \ldots, N$

$\quad\quad F[n,t] = P(z_t = \perp)F[n,t-1] + P(z_t = y_n)F[n-1,t-1]$

Back-Propagation

$$\mathcal{L} = -\ln F[N, T]$$

We can now back-propagate through this computation.

END