

TTIC 31230, Fundamentals of Deep Learning

David McAllester, Winter 2019

Connectionist Temporal Classification (CTC)

Connectionist Temporal Classification (CTC)

A Successful Deep Latent Variable Model

Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks

Alex Graves, Santiago Fernandez, Faustino Gomez, Jurgen Schmidhuber, ICML 2006

CTC

A speech signal $x[T, J]$ is labeled with a phone sequence $y[N]$ with $N \ll T$.

$x[t, J]$ is a speech signal vector.

$y[n] \in \mathcal{P}$ for a set of phonemes \mathcal{P} . 音位

The length N of $y[N]$ is not determined by T and the correspondence between n and t is not given.

$$\Phi^* = \underset{\Phi}{\operatorname{argmin}} E_{\langle x, y \rangle \sim \text{Train}} P_{\Phi}(y[N] \mid x[T, J]) \quad N \ll T$$

The CTC Model

We define a model

$$\underline{P_{\Phi}(z[T] \mid x[T, J])}$$

$$\underline{z[t] \in \mathcal{P} \cup \{\perp\}}$$

$y[N]$ is the result of removing \perp from $z[T]$.

$$z[T] \Rightarrow y[N]$$

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

Run a
RNN to calc
phone or blank

The CTC Model

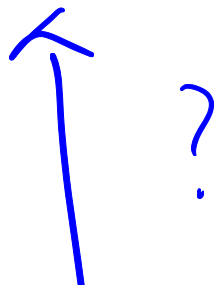
For $p \in \mathcal{P} \cup \{\perp\}$ we have an embedding vector $e[p, I]$. The embedding is a parameter of the model.

We take the phonemes $z[t]$ to be independently distributed.

$$p_{\Phi}(Z[T] \mid x[T, J]) = \prod_t P_{\Phi}(z[t] \mid x[T, J])$$

$$h[T, \tilde{J}] = \text{RNN}_{\Phi}(x[T, J])$$

$$\underbrace{P_{\Phi}(z[t] \mid x[T, J])}_{\text{?}} = \underset{z[t]}{\text{softmax}} \ e[z[t], I] \ W[I, \tilde{J}] \ h[t, \tilde{J}]$$



each z is independent

Dynamic Programming

Let $\vec{y}[t]$ to be the prefix of $y[N]$ emitted by the first t elements of z .

n : index of phone

t : index of time

$$\vec{y}[t] = z[1:t] - \perp$$

$$F[n, t] = P(\vec{y}[t] = y[1:n])$$

sequence of phone

probability at
time t you output

first n phones of
gold

$$F[0, 0] = 1$$

$$\text{For } n = 1, \dots, N \quad F[n, 0] = 0$$

$$\text{For } t = 1, \dots, T$$

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

$$\text{For } n = 1, \dots, N$$

$$F[n, t] = P(z[t] = \perp) F[n, t-1] + P(z[t] = y[n]) F[n-1, t-1]$$

$z[t]$ is independent

goal: at time t
I have $y[1:n]$

Back-Propagation

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

We can now back-propagate through this computation.

END