

CTC problem: we don't know the alignment of outputs with inputs, which is required by RNN, so CTC consider all the possible alignment and get their probability with DP.
The loss is just log (-probability), so we just need to max the probability.

TTIC 31230 Fundamentals of Deep Learning

Problems for CTC.

Problem 1. Dynamic Programing for HMMs Assume we have an input sequence x_1, \dots, x_T and a phoneme gold label y_1, \dots, y_T with $y_t \in \mathcal{P}$. This problem is simpler than CTC because the gold label has the same length as the input sequence.

In an HMM we assume a hidden state sequence s_1, \dots, s_T with $s_t \in \mathcal{S}$ where \mathcal{S} is some finite sets of “hidden states”. Here will assume that then some deep network has computed transition probabilities and emission probabilities.

$$P_{\text{Trans}}(s_{t+1} \mid s_t)$$

$$P_{\text{Emit}}(y_t \mid s_t)$$

We assume an initial state s_{init} and a stop state s_{stop} such that $s_1 = s_{\text{init}}$ (before emitting any phonemes). The length T is determined by when the hidden state becomes s_{stop} giving $s_{T+1} = s_{\text{stop}}$.

For a given gold sequence y_1, \dots, y_T we define a “forward tensor” as

$$F[t, s] = P(y_1, \dots, y_{t-1} \wedge s_t = s)$$

We have

$$\begin{aligned} F[1, s_{\text{init}}] &= 1 \\ F[1, s] &= 0 \quad \text{for } s \neq s_{\text{init}} \end{aligned}$$

(a) Write a dynamic programming equation to compute $F[t, s]$ from $F[t-1, s']$ for various values of s' .

Solution:

$$F[t, s] = \sum_{s'} F[t-1, s'] P_{\text{Emit}}(y_{t-1} \mid s') P_{\text{Trans}}(s \mid s')$$

(b) Express $P(y_1, \dots, y_T)$ in terms of $F[t, s]$.

Solution:

$$P(y_1, \dots, y_T) = F[T+1, s_{\text{stop}}]$$

(c) Explain why, if the forward equations are written in a framework, we do not need to also implement “backward” equations to compute

$$B[t, s] = P(y_t, \dots, y_T \mid s_t = s).$$

Solution: Once we have expressed the loss $-\ln P(y_1, \dots, y_T)$ in a framework we can train the model by SGD using the framework’s implementation of back-propagation. Nothing more is needed.

Problem 2. CTC for image labeling

Suppose that the training data consists of pairs (I, S) where I is an image and S is a set of object types occurring in the image. For example S might be $\{\text{Person, Dog, Car}\}$. To be concrete we can take \mathcal{C} to be the set of image labels used in CIFAR 100 and take S to be a subset of \mathcal{C} containing no more than five labels ($|S| \leq 5$). We want to do SGD on a model defining $P_\Phi(S \mid I)$.

We will use a latent variable $z[X, Y]$ such that for pixel coordinates (x, y) we have $z[x, y] \in \mathcal{C} \cup \{\perp\}$. For a given $z[X, Y]$ define $S(z[X, Y])$ to be the set of classes appearing in $z[X, Y]$, i.e., $S(z[X, Y]) = \{c \mid \exists x, y \ z(x, y) = c\}$. Here the “semantic segmentation” $Z[X, Y]$ is analogous to the phoneme sequence $z[T]$ in CTC. Unlike the CTC model, the label S is a set rather than a sequence.

We assume a CNN (with convolutions of stride 1 to preserve spatial dimensions) followed by a softmax at each pixel to get a probability $P_\Phi(z[x, y] = c)$ for each pixel location (x, y) and each $c \in \mathcal{C} \cup \{\perp\}$ and where each pixel location has an independent probability distribution over classes. To simplify notation we can reshape the pixel locations into a linear sequence and replace $z[X, Y]$ by $z[T]$ with $T = X \times Y$ so we have $z[0], z[1], \dots, z[T-1]$.

Define

$$S_t = \{c \in \mathcal{C} \mid \exists t' \leq t \ z[t'] = c\}$$

For $U \subseteq S$ define

$$F[U, t] = P(S_t = U)$$

Note that for $|S| \leq 5$ there are at most 32 possible values of U . Give dynamic programming equations defining $F[U, 0]$ and defining $F[U, t+1]$ in term of $F[U', t]$ for various U' .

Solution:

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

$$\text{For } U \text{ a nonempty subset of } S \ F[U, 0] = 0$$

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

$$\text{For } U \subseteq S$$

$$F[U, t] = P(z[t] = \perp)F[U, t-1] + \sum_{c \in U} P(z[t] = c)(F[U \setminus c, t-1] + F[U, t-1])$$