Connectionist Temporal Classification

TANG Zhiyuan, CSLT

11.23.2015

Problem representation(1)

Sequence to sequence learning

Unsegmented real-valued input stream -> Discrete label sequence

Problem representation(1)

Sequence to sequence learning

Unsegmented real-valued input stream -> Discrete label sequence

-- Classification

-- Temporal Classification

-- Connectionist Temporal Classification, as **loss function**

Problem representation(2)

Input: $X = [x_1, x_2, \ldots, x_T]$, $x_n \in R^m$

Output: $Z = [z_1, z_2, \ldots, z_U]$, $z_m \in L$, L is a set of finite labels

$h(X) = Z$

-- input: unsegmented vs. discrete

Morse : 122 1 -> we(discrete, segmented)

Speech: waveform of /lʌvˈnɒlɪdʒ/ -> love knowledge(discrete, unseg)

Basic network settings for classification

Input

Output

FNN/CNN/RNN

LSTM(Long short-term memory)

CTC solution taking ASR as an example

Input: $X = [x_1, x_2, \dots, x_T]$, $x_T \in R^m$, x_T is one **frame** of speech feature

Output: $Z = [z_1, z_2, \dots, z_U]$, $z_U \in L$, L is a set of finite **phone** labels

DNN: $h(X) = Z$

-- T vs. U, at least in Speech Recognition

$U \leq T$

uncrossed

CTC solution taking ASR as an example

Input: $X = [x_1, x_2, \ldots , x_T]$, $x_T \in R^m$, x_T is one **frame** of speech feature

Output: $Z = [z_1, z_2, \ldots , z_U]$, $z_U \in L$, L is a set of finite **phone** labels

DNN: $h(X) = Z$

CTC solution taking ASR as an example

Input: $X = [x_1, x_2, \ldots, x_T]$, $x_T \in R^m$, x_T is one **frame** of speech feature

Output: $Z = [z_1, z_2, \ldots, z_U]$, $z_U \in L$, L is a set of finite **phone** labels

DNN: $f(X) = \pi$, π is a label sequence of length T

Map: $ß(\pi) = \pi^* = Z$, egs, $ß(a - ab-) = ß(-aa - -abb) = aab$

CTC solution taking ASR as an example

Input: $X = [x_1, x_2, \ldots, x_T]$, $x_T \in R^m$, x_T is one **frame** of speech feature

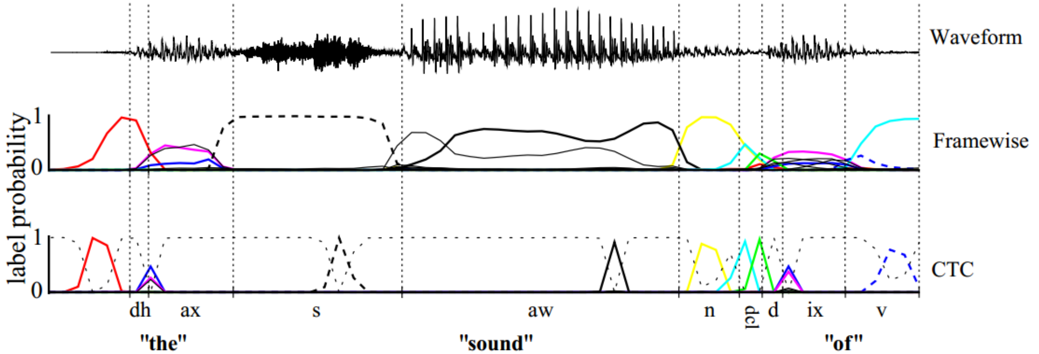
Output: $Z = [z_1, z_2, \ldots, z_U]$, $z_U \in L$, L is a set of finite **phone** labels

DNN: $f(X) = \pi$, π is a label sequence of length T

Map: $ß(\pi) = \pi^* = Z$, egs, $ß(a - ab-) = ß(-aa - -abb) = aab$

-- The symbol ”-” means nothing is output, that is, **blank**.

Doesn’t mean it doesn’t contribute, but accumulating, then **spike**.



Details(formulas mostly)

Propagate

-- probability to generate **any** sequence π

$$p(\pi|\mathbf{x}) = \prod_{t=1}^{T} y_{\pi_t}^{t}, \ \forall \pi \in {L'}^{T}$$

Details(formulas mostly)

Propagate

-- probability to generate **any** sequence π

$$p(\pi|\mathbf{x}) = \prod_{t=1}^{T} y_{\pi_t}^{t}, \ \forall \pi \in {L'}^{T}$$

-- probability to generate a **specific** sequence **l,** to be **maximized**

$$p(\mathbf{l}|\mathbf{x}) = \sum_{\pi \in \mathcal{B}^{-1}(\mathbf{l})} p(\pi|\mathbf{x})$$

Details(formulas mostly)

Propagate

-- probability to generate **any** sequence π

$$p(\pi|\mathbf{x}) = \prod_{t=1}^{T} y_{\pi_t}^{t}, \ \forall \pi \in {L'}^{T}$$

-- probability to generate a **specific** sequence **l,** to be **maximized**

$$p(\mathbf{l}|\mathbf{x}) = \sum_{\pi \in \mathcal{B}^{-1}(\mathbf{l})} p(\pi|\mathbf{x})$$

Feedforward

-- select **some** sequence with biggest probability

$$h(\mathbf{x}) = \arg \max_{\mathbf{l} \in L^{\leq T}} p(\mathbf{l}|\mathbf{x})$$

Details(formulas mostly)

Backpropagate

-- probability to generate a **specific** sequence **l,** to be **maximized**

$$p(\mathbf{l}|\mathbf{x}) = \sum_{\pi \in \mathcal{B}^{-1}(\mathbf{l})} p(\pi|\mathbf{x})$$

Details(formulas mostly)

Backpropagate

-- calculate it first, using **Forward-Backward** algorithm

$$p(\mathbf{l}|\mathbf{x}) = \sum_{\pi \in \mathcal{B}^{-1}(\mathbf{l})} p(\pi|\mathbf{x})$$

-- maximize likelihood

minimize -log of it, using **Gradient Descent** algorithm

$$O^{ML}(S, \mathcal{N}_w) = -\sum_{(\mathbf{x},\mathbf{z}) \in S} ln\big(p(\mathbf{z}|\mathbf{x})\big)$$

Finally

select **some** sequence with biggest probability

-- search problem, using **Prefix search** etc.

$$h(\mathbf{x}) = \arg\max_{\mathbf{l} \in L^{\leq T}} p(\mathbf{l}|\mathbf{x})$$

Reference

Graves, Alex, et al. "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks." *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006.

Thanks.