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

Коннекционистская временная классификация

Руслан Ахтариев
НИУ ВШЭ

- Рукописные тексты
- Речь

- Большое количество шума
- Нет сегментации
- Пространство ответов сильно меньше пространства входных данных

Basics

$$LED(h, S') = \frac{1}{|S'|} \sum_{(x,z) \in S'} ED(h(x))$$

$ED(p, q)$ – **edit distance**

$x = (x_1, x_2, \ldots, x_T)$ – **вход**, $x_i \in \mathbb{R}$

$z = (z_1, z_2, \ldots, z_U)$ – **выход**, $z_i \in L, |L| < \infty, U \leqslant T$

y_k^t – **вероятность класса** k **в момент** $t \implies$

$\implies$ **распределение над последовательностями длины** T **из алфавита** $L' = L \cup \{''\}$

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

$$B(a - ab-) = B(-aa - -abb-) = aab$$

$l \in L^{\leqslant T}$ – **вектор итоговых классов**

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

Classifier

Perfect

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

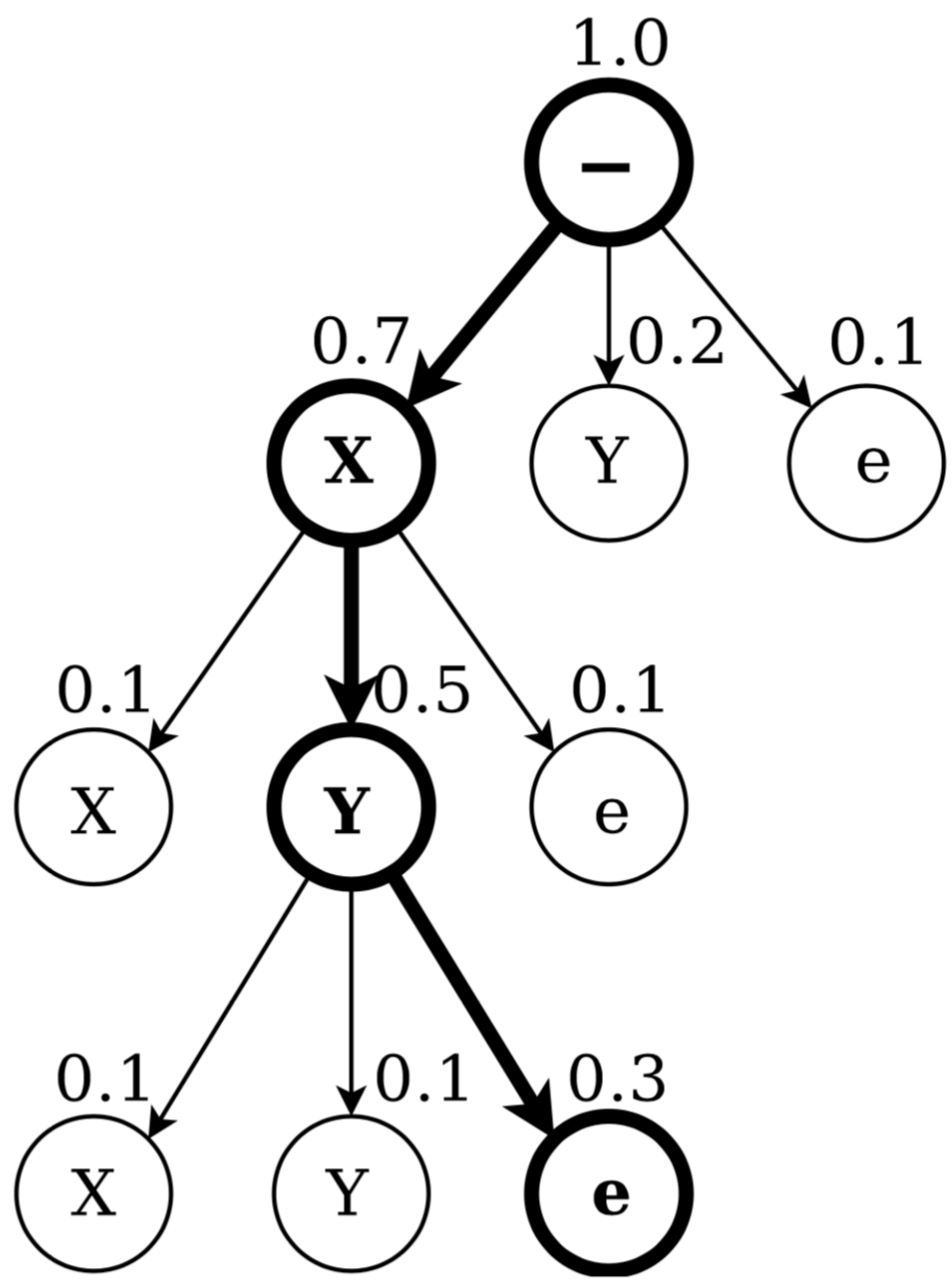
Best path decoding

$$\pi^* = arg \max_{\pi \in N^t} p(\pi|x)$$

$$h(x) \approx B(\pi^*)$$

Просто конкатенация самых вероятных классов в каждый момент

Prefix search decoding



Forward

$\alpha_t(s)$ – **вероятность префикса** l **длины** s **в момент** t

$$\alpha_t(s) = \sum_{\pi \in N^T; B(\pi_{1:t}) = l_{1:s}} \prod_{t'=1}^{t} y^{t'}_{\pi_{t'}}$$

$$l' = -l_1 - l_2 - l_3 - \ldots - l_{|l|} -$$

Переходы:

- Пустая строк -> символ
- Символ -> другой символ

$$\alpha_1(1) = y^1_b$$

$$\alpha_1(2) = y^1_{l_1}$$

$$\alpha_1(s) = 0, \forall s > 2$$

$$\alpha_t(s) = \begin{cases} \bar{\alpha}_t(s) y^t_{\mathbf{l}'_s} & \text{if } \mathbf{l}'_s = b \text{ or } \mathbf{l}'_{s-2} = \mathbf{l}'_s \\ \big(\bar{\alpha}_t(s) + \alpha_{t-1}(s-2)\big) y^t_{\mathbf{l}'_s} & \text{otherwise} \end{cases}$$

$$\bar{\alpha}_t(s) \stackrel{\text{def}}{=} \alpha_{t-1}(s) + \alpha_{t-1}(s-1).$$

$$p(l \mid x) = \alpha_T(|l'|) + \alpha_T(|l'| - 1)$$

Backward

$$\beta_t(s) = \sum_{\pi \in N^T; B(\pi_{t:T}) = l_{s:|l|}} \prod_{t'=1}^{t} y_{\pi_{t'}}^{t'}$$

$$\beta_T(|l'|) = y_b^T$$

$$\beta_T(|l'| - 1) = y_{l_{|l|}}^T$$

$$\beta_T(s) = 0, \forall s < |l'| - 1$$

$$\beta_t(s) = \begin{cases} \bar{\beta}_t(s) y_{\mathbf{l}'_s}^t & \text{if } \mathbf{l}'_s = b \text{ or } \mathbf{l}'_{s+2} = \mathbf{l}'_s \\ \left(\bar{\beta}_t(s) + \beta_{t+1}(s+2)\right) y_{\mathbf{l}'_s}^t & \text{otherwise} \end{cases}$$

$$\bar{\beta}_t(s) \stackrel{\text{def}}{=} \beta_{t+1}(s) + \beta_{t+1}(s+1).$$

Rescaling

$$C_t = \sum_s \alpha_t(s) \qquad \hat{\alpha}_t(s) = \frac{\alpha_t(s)}{C_t}$$

$$D_t = \sum_s \beta_t(s) \qquad \hat{\beta}_t(s) = \frac{\beta_t(s)}{D_t}$$

$$\ln(p(l\,|\,x)) = \sum_{t=1}^{T} \ln(C_t)$$

Maximum likelihood

$$O(S,Y) = -\sum_{(x,z)\in S} \ln(p(z|x)) \rightarrow min$$

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

$$\frac{\partial O(\{(x,z)\},Y)}{\partial y_k^t} = -\frac{\partial \ln(p(z|x))}{\partial y_k^t}$$

$$\alpha_t(s)\beta_t(s) = \sum_{\pi \in B^{-1}(l);\pi^t = l'_s} y_{l'_s}^t \prod_{t=1}^{T} y_{\pi_t}^t$$

$$p(l|x) = \sum_{s=1}^{|l'|} \frac{\alpha_t(s)\beta_t(s)}{y_{l'_s}^t}$$

$$\frac{\alpha_t(s)\beta_t(s)}{y_{l'_s}^t} = \sum_{\pi \in B^{-1}(l);\pi^t = l'_s} p(\pi|x)$$

$$lab(l,k) = \{s : l_s = k\}$$

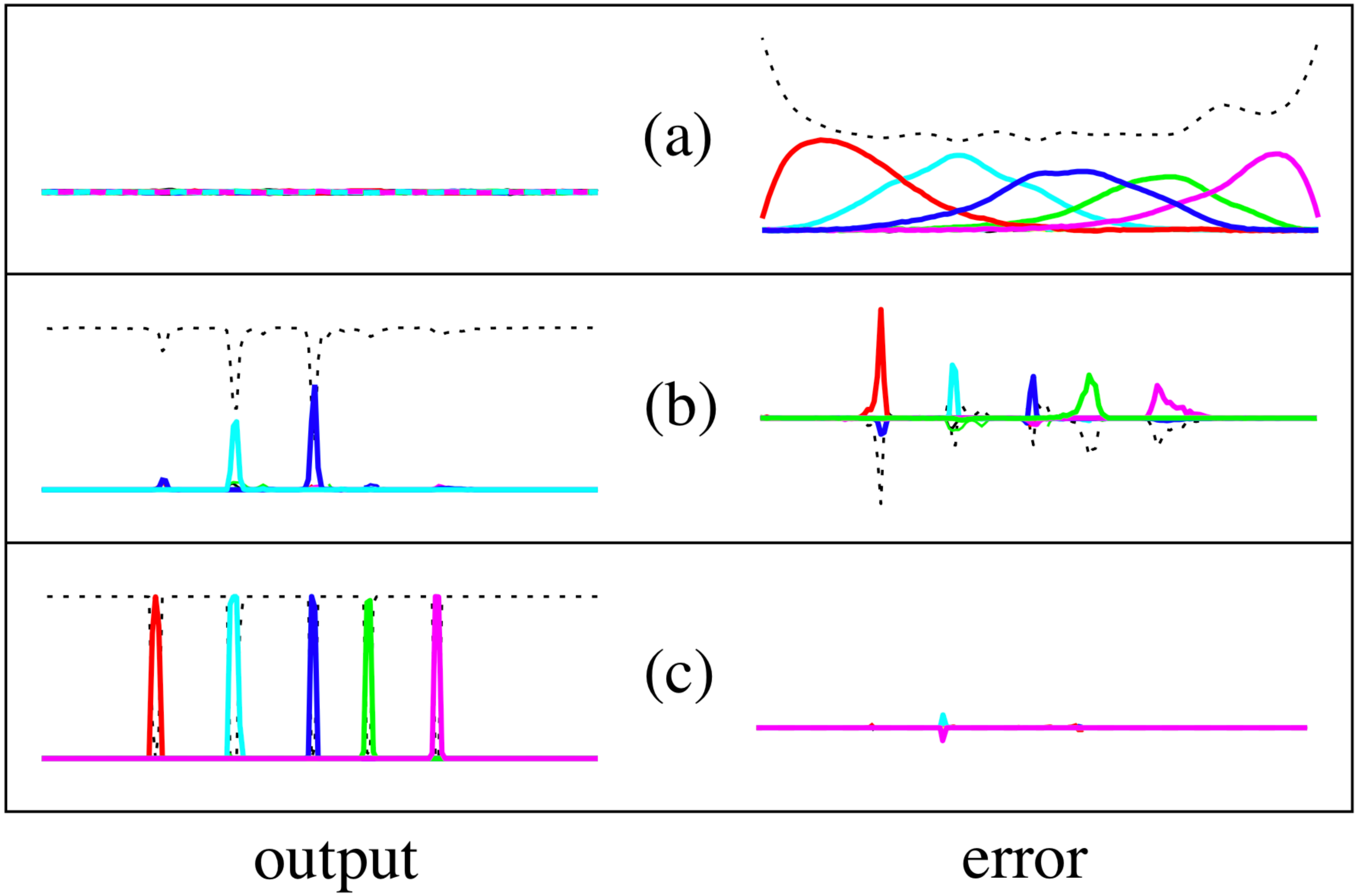
$$\frac{\partial p(l|x)}{\partial y_k^t} = \frac{1}{y_k^{t^2}} \sum_{s \in lab(l,k)} \alpha_t(s)\beta_t(s)$$

$$\frac{\partial O(\{(x,z)\},Y)}{\partial y_k^t} = \frac{1}{p(l|x)} \frac{\partial p(l|x)}{\partial y_k^t} = \frac{1}{\alpha_T(|l'|) + \alpha_T(|l'|-1)} \frac{1}{y_k^{t^2}} \sum_{s \in lab(l,k)} \alpha_t(s)\beta_t(s)$$

Rescaling

$$\frac{\partial O(\{(x,z)\},Y)}{\partial u_k^t} = y_k^t - \frac{1}{y_k^t Z_t} \sum_{s \in lab(z,k)} \hat{\alpha}_t(s)\hat{\beta}_t(s)$$

$$Z_t = \sum_{s=1}^{|l'|} \frac{\hat{\alpha}_t(s)\hat{\beta}_t(s)}{y_{l'_s}^t}$$


(a)
(b)
(c)
output
error

Experiments

HMM — Hidden Markov Models

System	LER
Context-independent HMM	38.85 %
Context-dependent HMM	35.21 %
BLSTM/HMM	33.84 ± 0.06 %
Weighted error BLSTM/HMM	31.57 ± 0.06 %
CTC (best path)	31.47 ± 0.21 %
CTC (prefix search)	30.51 ± 0.19 %