

TME 5 (AMAL) → RNN (TME 4) / ① LSTM/GRU

② Generation

LSTM

$$h_1 \xrightarrow{A^T} h_2 \xrightarrow{A^T} \dots \xrightarrow{A^T} h_N$$

$$h_N = (A^T)^N h_1$$

Porte →  $\sigma$  (transp. linéaire)  $\in [0, 1]$

Mémoire interne  $C_t \in \mathbb{R}^n$  ( $h_t \in \mathbb{R}^n$ )

À chaque pas de temps

① Mettre à jour la mémoire interne

$$C_{t+1} = G_f(x_t, h_t) * C_t + G_i(x_t, h_t) * \gamma_i(x_t, h_t)$$

produit terme à terme (knocker)

$h_t$  est utilisé pour faire des prédictions

② Calculer l'état  $h_{t+1}$

$$h_{t+1} = G_o(x_t, h_t) * \tanh(C_{t+1})$$

G

Séquences de taille variable

beam-search

$$\frac{\partial L}{\partial h_1} = A^N \frac{\partial L}{\partial h_N}$$

$$A = U \Sigma U^T$$

$$A^N = U \Sigma^N U^T$$

il est important de surveiller le gradient

explosion du gradient

gradient clipping

annulation du gradient

LSTM/GRU résout ce pb

## Séquences de taille variable

$a_1, a_2, a_3$

$b_1, b_2$

$c_1$

→ batch →

$a_1$

$b_1$

$c_1$

$a_2$

$b_2$

[PAD]

$a_3$

[PAD]

[PAD]

Dataset → renvoie une séquence (de taille variable) ex  $[b_1, b_2]$

Dataloader (...., collate\_fn = f)

f (liste de liste d'entiers) → batch

ex f:  $[[a_1, a_2, a_3], [b_1, b_2], [c_1]]$

① Calculer la taille max

② Renvoyer un tenseur (LongTensor) en utilisant [PAD]



$x =$

$a_1$	$b_1$	$c_1$
$a_2$	$b_2$	$\emptyset$
$a_3$	$\emptyset$	$\emptyset$

[PAD]

forward

$h_1^a$	$h_1^b$	$h_1^c$
$h_2^a$	$h_2^b$	$h_2^c$
$h_3^a$	$h_3^b$	$h_3^c$

decoder

$y_1^a$	$y_1^b$	$\dots$
	$\dots$	$\dots$
	$\emptyset$	$\emptyset$

CE loss

$x == \text{PAD\_ix}$

somme  
 $\div 6$

$l_1^a$	$l_1^b$	$l_1^c$
$l_2^a$	$l_2^b$	$0$
$l_2^a$	$0$	$0$

somme

1	1	1
1	1	0
1	0	0

masque  
qui correspond  
au batch

$l_1^a$	$l_1^b$	$l_1^c$
$l_2^a$	$\dots$	$\dots$
$l_3^a$	$\dots$	$l_3^c$



Comment savoir qu'une phrase se termine?

$[a_1, a_2, a_3] \rightarrow [a_1, a_2, a_3, [EOS]]$  ) au niveau de Dataset.

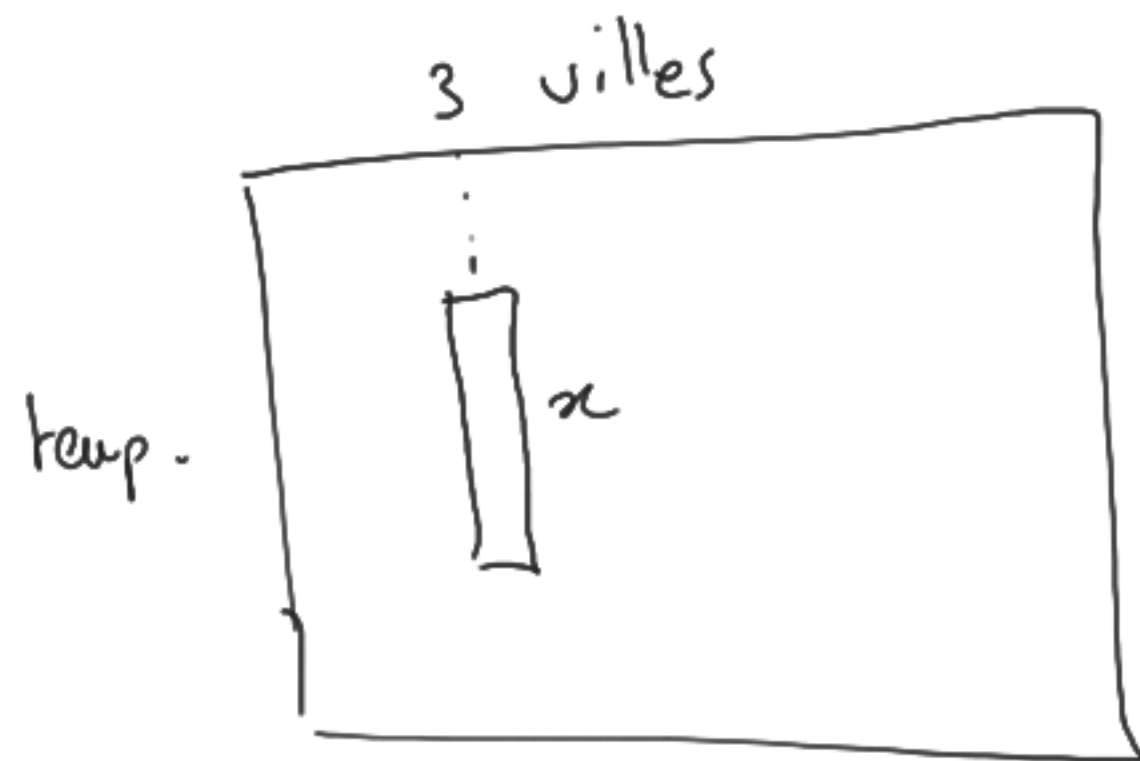


TP 4 = one-hot  $\Rightarrow$  TP 5 = nm. Embedding  $(N, n) = \text{emb}$

↓  
associe à chaque  
entier entre  $[0, N[$  un  
vecteur dans  $\mathbb{R}^n$

Pour

$$\underbrace{[0, N[}_{\mathcal{X}}^{T \times B} \rightarrow \underbrace{\mathbb{R}^{T \times B \times n}}_{\text{emb}(\mathcal{X})}$$



• Dataloader

--getitem--

$(x, 3)$

• DataLoader

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_D \end{pmatrix} / \begin{pmatrix} 3 \\ \vdots \\ \vdots \\ \vdots \end{pmatrix}$$