COMPRESSING WORD EMBEDDINGS VIA DEEP COMPOSITIONAL CODE LEARNING

Language Reading Group

February 7, 2019

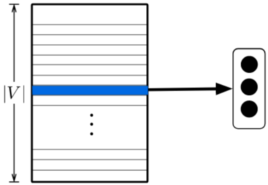
Raphael Shu and Hideki Nakayama

University of Tokyo

Normalement

mot	$\rightarrow$	$[1, V]$
penguin	$\mapsto$	42

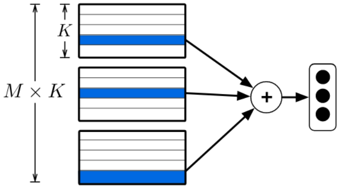
$$\tilde{E}(w) = L_w$$



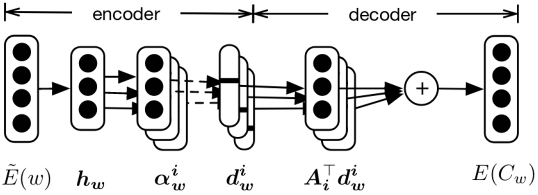
Le papier

mot	$\rightarrow$	$[1, K]^M$
penguin	$\mapsto$	0 7 0 1 7 3 6 0

$$E(C_w) = \sum_{i=1}^{M} L^i_{C^i_w}$$



$$(\hat{C}, \hat{E}) = \underset{C,E}{\operatorname{argmin}} \frac{1}{|V|} \sum_{w \in V} ||E(C_w) - \tilde{E}(w)||^2$$



$$E(C_w) = \sum_{i=0}^{M} \boldsymbol{A_i}^\top \boldsymbol{d_w^i}$$

$$\left(\boldsymbol{d_w^i}\right)_k = \mathrm{softmax}\ \left(\log \boldsymbol{\alpha_w^i} + G\right)_k$$

$$\boldsymbol{\alpha_w^i} = \mathrm{softplus}(\boldsymbol{\theta_i'}^\top \boldsymbol{h_w} + \boldsymbol{b_i'})$$

$$\boldsymbol{h_w} = \tanh(\boldsymbol{\theta}^\top \tilde{E}(w) + \boldsymbol{b})$$

$$C_w^i = \operatorname*{argmax}_k \left((d_w^i)_k\right)$$

category	word	8×8 **code**	16×16 **code**
animal	dog	0 7 0 1 7 3 7 0	7 7 0 8 3 5 8 5 B 2 E E 0 B 0 A
	cat	7 7 0 1 7 3 7 0	7 7 2 8 B 5 8 C B 2 E E 4 B 0 A
	penguin	0 7 0 1 7 3 6 0	7 7 E 8 7 6 4 C F D E 3 D 8 0 A
verb	go	7 7 0 6 4 3 3 0	2 C C 8 2 C 1 1 B D 0 E 0 B 5 8
	went	4 0 7 6 4 3 2 0	B C C 6 B C 7 5 B 8 6 E 0 D 0 4
	gone	7 7 0 6 4 3 3 0	2 C C 8 0 B 1 5 B D 6 E 0 2 5 A

	#vectors	**vector size**	**code len**	**code size**	**total size**	**accuracy**
GloVe baseline	75102	78 MB	-	-	78 MB	87.18
prune 80%	75102	21 MB	-	-	21 MB	86.25
prune 90%	75102	11 MB	-	-	11 MB	84.96
NPQ (10×256)	256	0.26 MB	80 bits	0.71 MB	0.97 MB	86.21
NPQ (60×256)	256	0.26 MB	480 bits	4.26 MB	4.52 MB	87.11
8×64 coding	512	0.52 MB	48 bits	0.42 MB	0.94 MB	86.66
16×32 coding	512	0.52 MB	80 bits	0.71 MB	**1.23 MB**	87.37
32×16 coding	512	0.52 MB	128 bits	1.14 MB	1.66 MB	87.80
64×8 coding	512	0.52 MB	192 bits	1.71 MB	2.23 MB	**88.15**