Word representations

PSTALN

Word2Vec

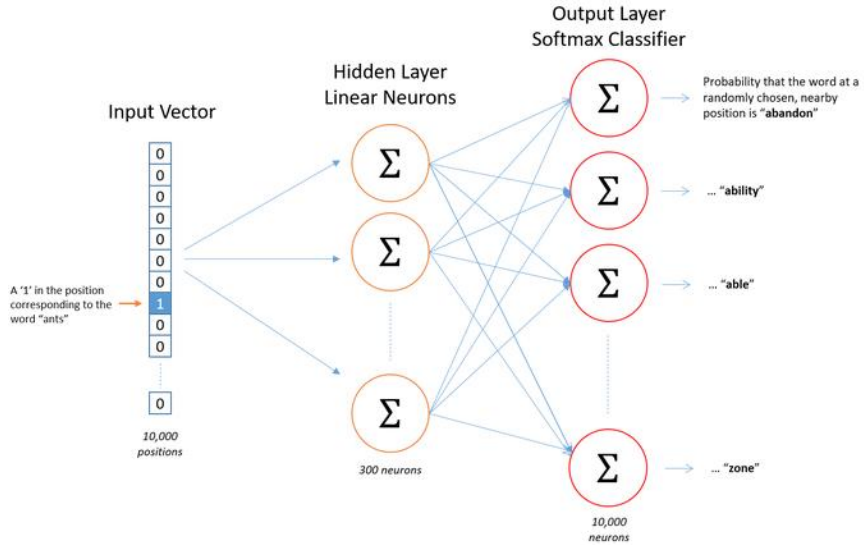
Le contexte d'un mot:

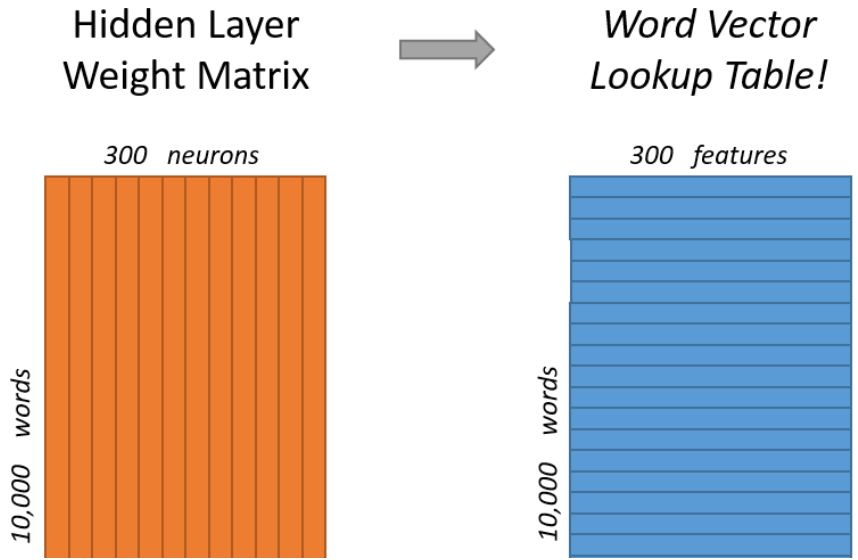
The | quick | brown | **fox** | jumps | over | the lazy dog. ⟶

(X , Y)

(fox, quick)
(fox, brown)
(fox, jumps)
(fox, over)

Word2Vec (Skipgram et CBOW)

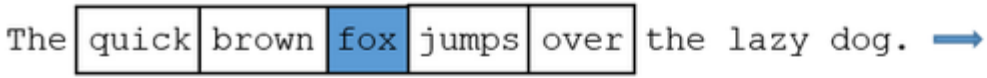




Source: http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model/

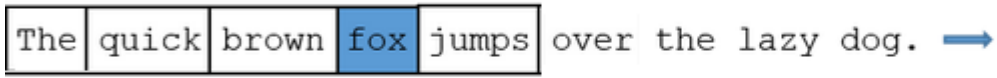
Word2Vec

POS



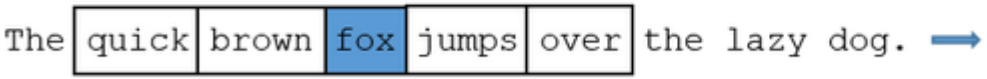
(X , Y)
(fox, quick +ADJ)
(fox, brown +ADJ)
(fox, jumps +V)
(fox, over +...)

Dependencies



(X , Y)
(fox, quick +mod)
(fox, brown +mod)
(fox, jumps -subj)
(fox, The + det)

Position



(X , Y)
(fox, quick -2)
(fox, brown -1)
(fox, jumps +1)
(fox, over +2)

Contextualized Word Representations (ELMo)

Each word representation should be a function of its context

- **Solves Polysemy**

How?

- **Train a Neural Language model**

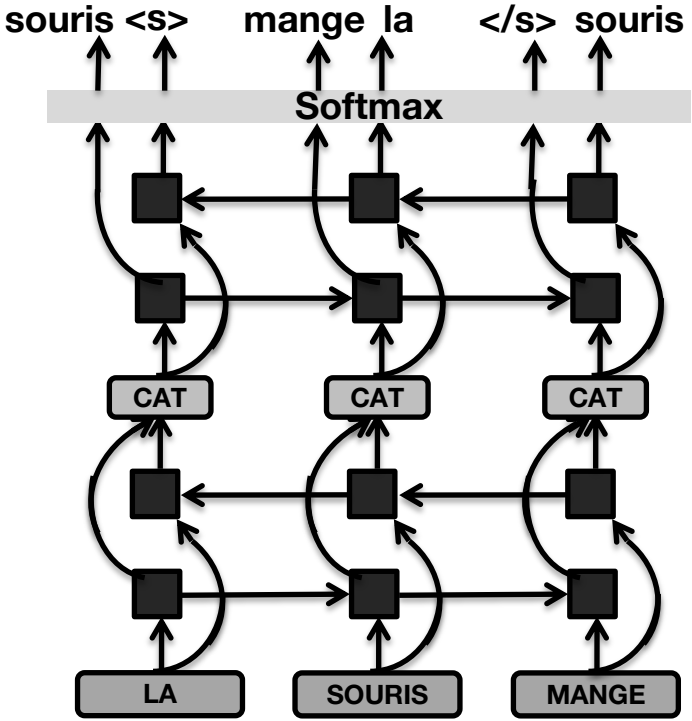
$$p(t_1, t_2, \ldots, t_N) = \prod_{k=1}^{N} p(t_k \mid t_{k+1}, t_{k+2}, \ldots, t_N)$$

$$p(t_1, t_2, \ldots, t_N) = \prod_{k=1}^{N} p(t_k \mid t_1, t_2, \ldots, t_{k-1}).$$

- **Use its hidden states as word embeddings**

$$R_k = \{\mathbf{x}_k^{LM}, \overrightarrow{\mathbf{h}}_{k,j}^{LM}, \overleftarrow{\mathbf{h}}_{k,j}^{LM} \mid j = 1, \ldots, L\}$$

$$\mathbf{ELMo}_k^{task} = E(R_k; \Theta^{task}) = \gamma^{task} \sum_{j=0}^{L} s_j^{task} \mathbf{h}_{k,j}^{LM}.$$



Contextualized Word Representations (ELMo)

Task	Previous SOTA		Our baseline	ELMo + Baseline	Increase (Absolute/Relative)
SQuAD	SAN	84.4	81.1	85.8	4.7 / 24.9%
SNLI	Chen et al (2017)	88.6	88.0	88.7 +/- 0.17	0.7 / 5.8%
SRL	He et al (2017)	81.7	81.4	84.6	3.2 / 17.2%
Coref	Lee et al (2017)	67.2	67.2	70.4	3.2 / 9.8%
NER	Peters et al (2017)	91.93 +/- 0.19	90.15	92.22 +/- 0.10	2.06 / 21%
Sentiment (5-class)	McCann et al (2017)	53.7	51.4	54.7 +/- 0.5	3.3 / 6.8%

Source: https://allennlp.org/elmo