LSTM versus Markov chain in text generation

ATELIER ML ALLOHA 14/11/2018

Atelier Machine Learning Brest

ALLOHa | Advanced Living Lab For Ocean Hacks

Cf Anouck HUBERT @ PNBI

anouck.hubert@tech-brest-iroise.fr

Lien ALLOHa

Données, présentations & notebooks:

https://amlb.github.io

Speaker

Erwan KERIBIN

Brestois de naissance

Ecole d’ingé (ENSIIE - Evry) + master d’IA (UPMC - Paris)

Freelance Dev depuis fin 2017

Inspiration

Generating Text with Recurrent Neural Networks

Ilya Sutskever, James Martens, and Geoffrey Hinton, ICML 2011

Dataset wikipedia: 86-character alphabet of about 100MB

Training time: five days on 8 high-end GPUs

Result:

The meaning of life is the tradition of the ancient human reproduction: it is less favorable to the good boy for when to remove her bigger. In the show's agreement unanimously resurfaced. The wild pasteured with consistent street forests were incorporated by the 15th century BE. In 1996 the primary rapford undergoes an effort that the reserve conditioning, written into Jewish cities, sleepers to incorporate the .St Eurasia that activates the population. Mar??a Nationale, Kelli, Zedlat-Dukastoe, Florendon, Ptu's thought is. To adapt in most parts of North America, the dynamic fairy Dan please believes, the free speech are much related to the

Le dataset

https://www.legifrance.gouv.fr

72 codes

~120Mo de pdfs

70Mo+ en format txt

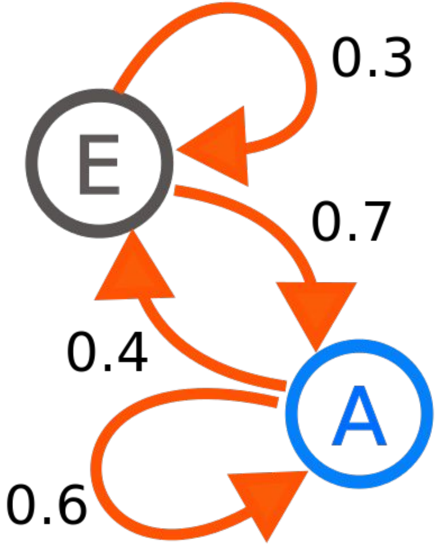
Chaîne de Markov

https://en.wikipedia.org/wiki/Markov_chain

Suite de probas

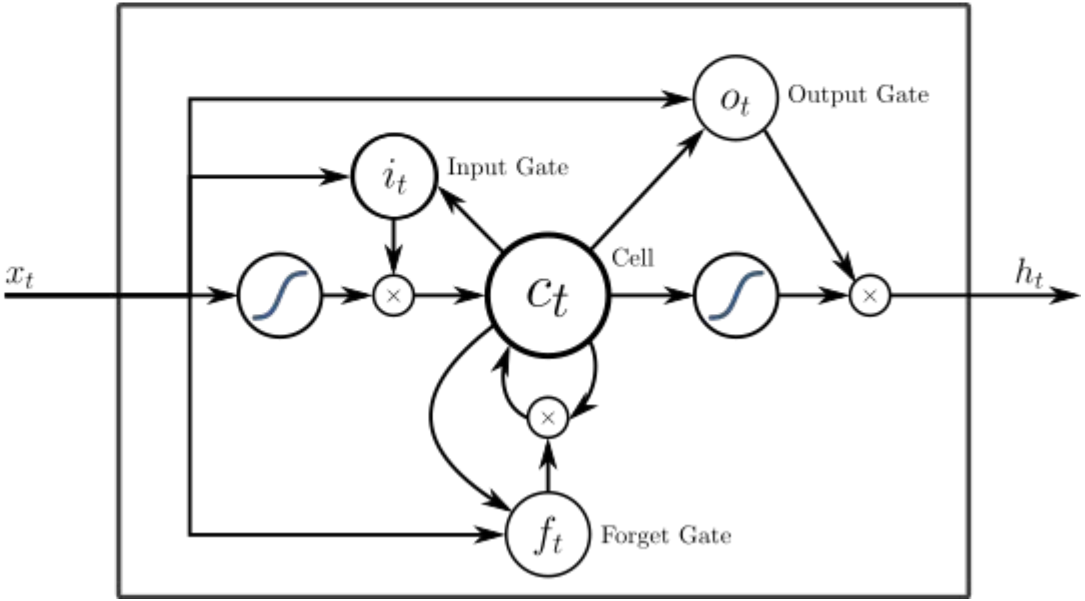
Probas d’enchaînements de mots

Probas d’enchaînements de charactères



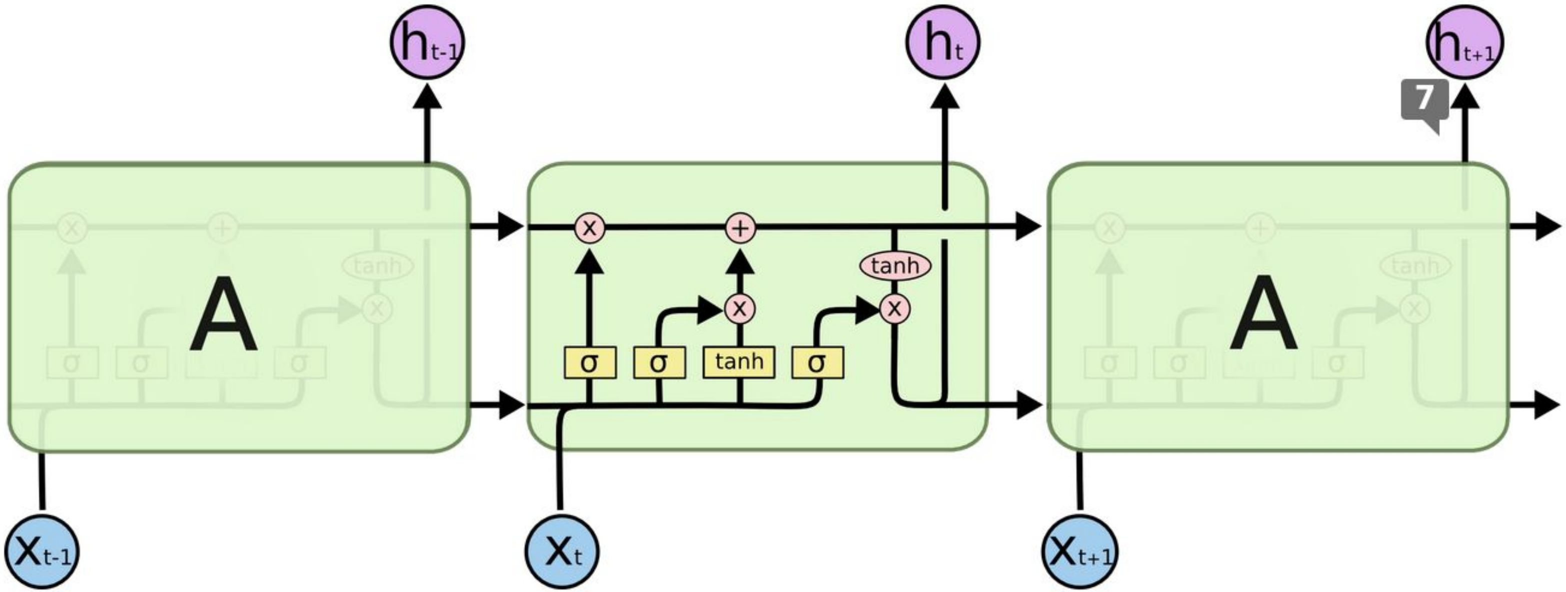
Long Short Term Memory

https://en.wikipedia.org/wiki/Long_short-term_memory



LSTM

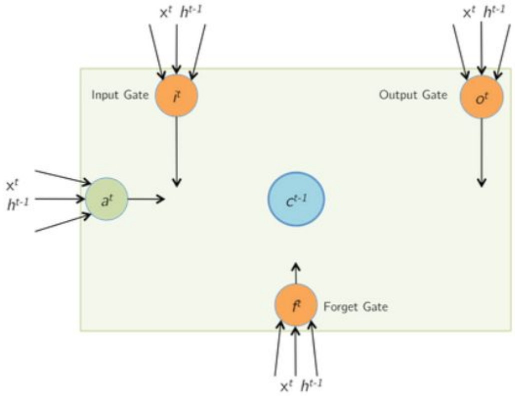
https://docs.google.com/presentation/d/1YGrpTPtUEZ7Z3rvgaRvsQFwJbqDnwrCFswVVJJjfY70/edit#slide=id.p



LSTM: http://arunmallya.github.io/writeups/nn/lstm

Forward Pass: Input and Gate Computation

At time t, The LSTM receives a new input vector x^t (including the bias term), as well as a vector of its output at the previous timestep, h^{t-1}.



$$
\begin{aligned}
a^t &= \tanh(W_c x^t + U_c h^{t-1}) = \tanh(\hat{a}^t) \\
i^t &= \sigma(W_i x^t + U_i h^{t-1}) = \sigma(\hat{i}^t) \\
f^t &= \sigma(W_f x^t + U_f h^{t-1}) = \sigma(\hat{f}^t) \\
o^t &= \sigma(W_o x^t + U_o h^{t-1}) = \sigma(\hat{o}^t)
\end{aligned}
$$

Ignoring the non-linearities,

$$
z^t = \begin{bmatrix} \hat{a}^t \\ \hat{i}^t \\ \hat{f}^t \\ \hat{o}^t \end{bmatrix} = \begin{bmatrix} W^c & U^c \\ W^i & U^i \\ W^f & U^f \\ W^o & U^o \end{bmatrix} \times \begin{bmatrix} x^t \\ h^{t-1} \end{bmatrix}
$$
$$
= W \times I^t
$$

If the input x^t is of size $n \times 1$, and we have d memory cells, then the size of each of W_* and U_* is $d \times n$, and and $d \times d$ resp. The size of W will then be $4d \times (n + d)$. Note that each one of the d memory cells has its own weights W_* and U_*, and that the only time memory cell values are shared with other LSTM units is during the product with U_*.