

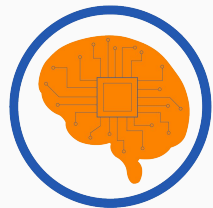
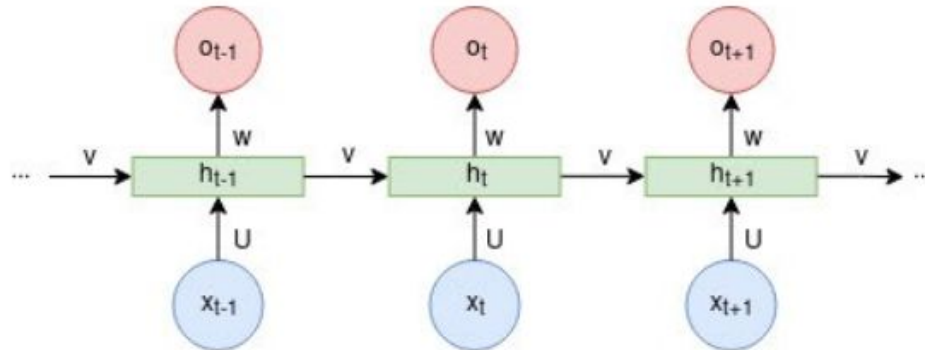
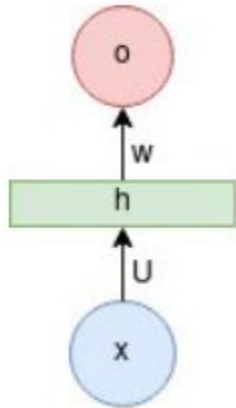
Elmo



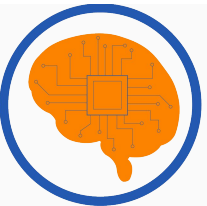
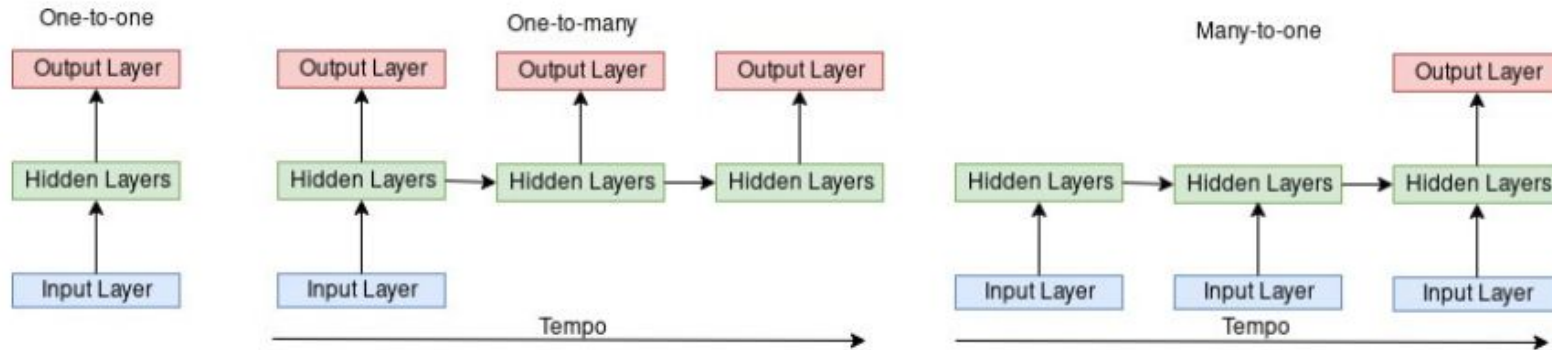
RNN/LSTM

RNN:

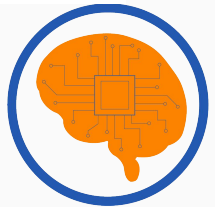
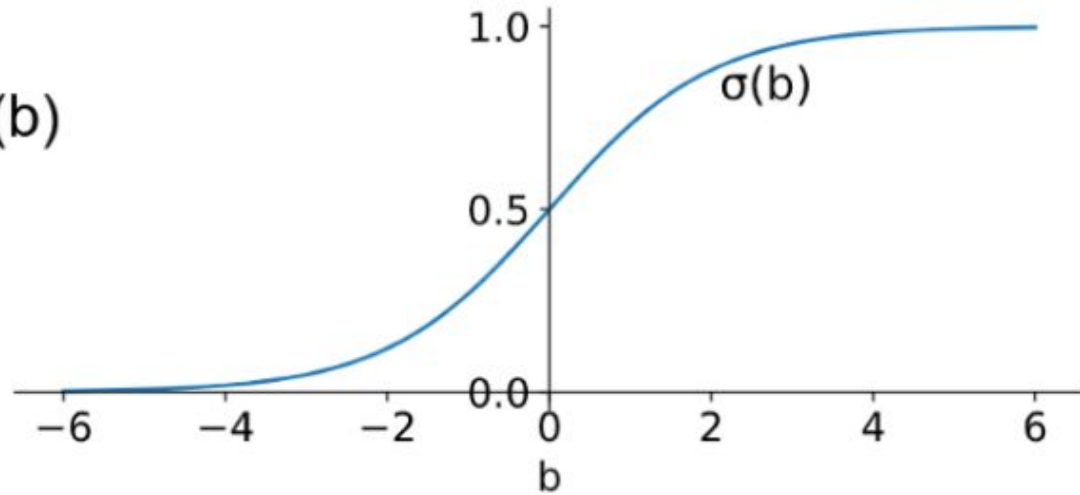
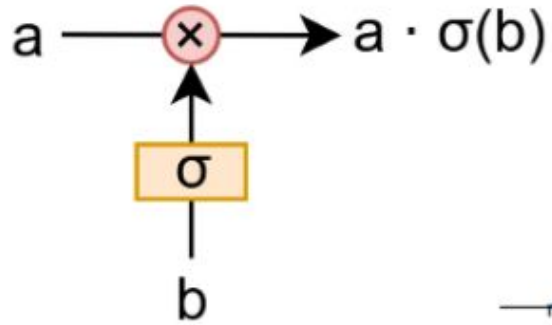
-Redes com memoria



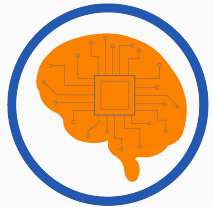
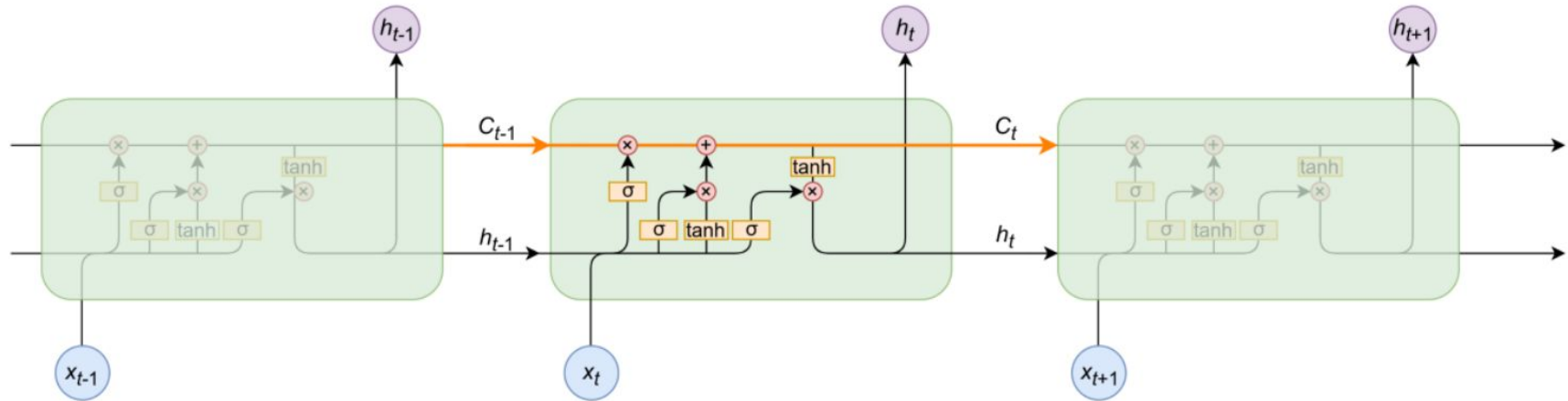
RNN/LSTM



RNN/LSTM



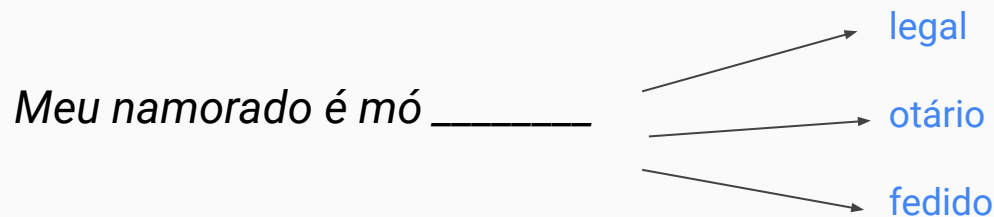
RNN/LSTM



Modelo de Língua

O que é um modelo de língua?

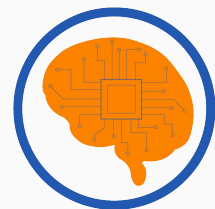
Um modelo de língua procura tentar prever qual é a próxima palavra em uma frase



Uma definição um pouco mais formal ...

Dada uma sequência de palavras $x(1), x(2), \dots, x(t)$, qual é a distribuição de probabilidade da próxima palavra $x(t+1)$?

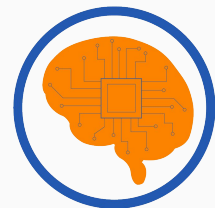
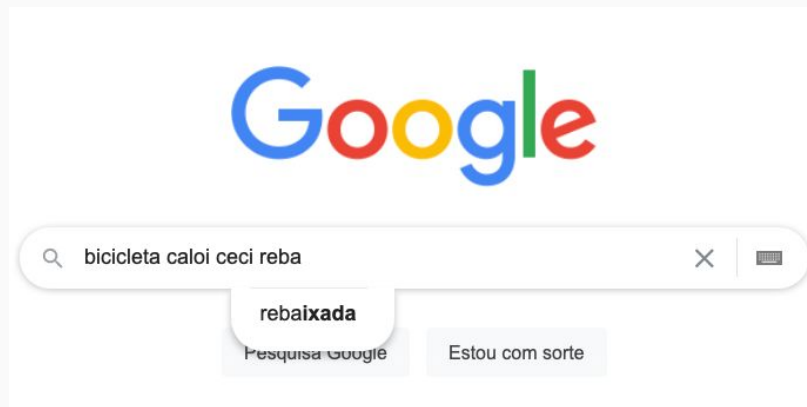
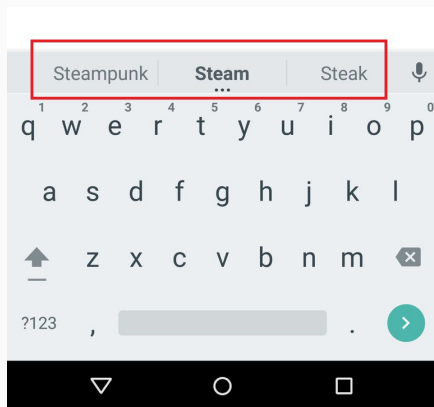
$$P(x(t+1) \mid x(t), \dots, x(1))$$



Modelo de Língua

Isso é o que chamamos de um modelo de língua!

Você também pode pensar nisso como um sistema que atribui uma probabilidade para uma porção de texto

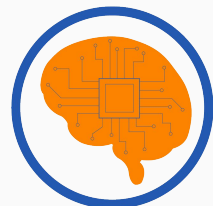
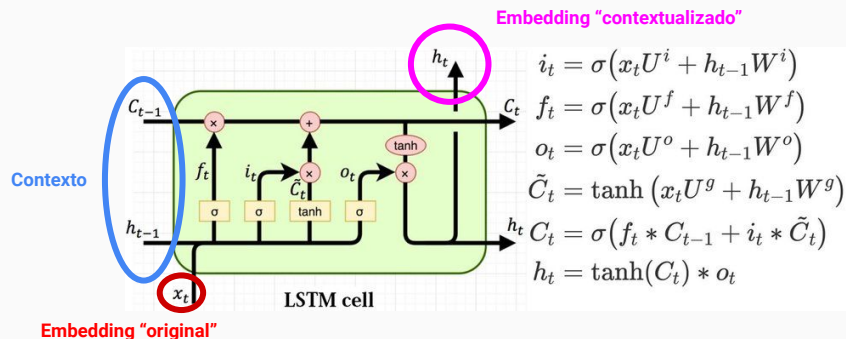


Elmo - introduzindo

*“Unlike most widely used word embeddings ELMo word representations are functions of the **entire input** sentence”*

ELMo = acrônimo de *Embeddings from Language Models*

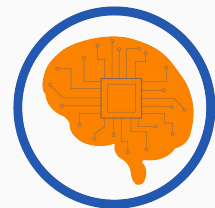
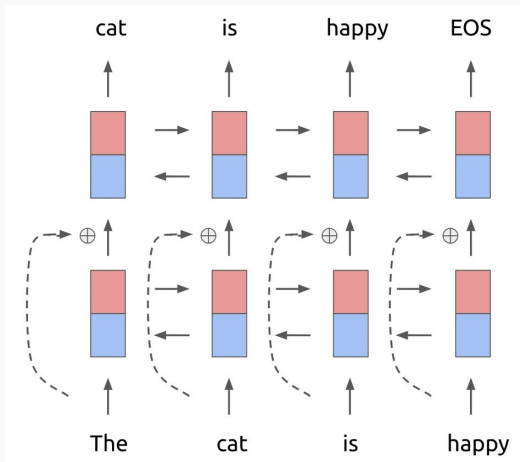
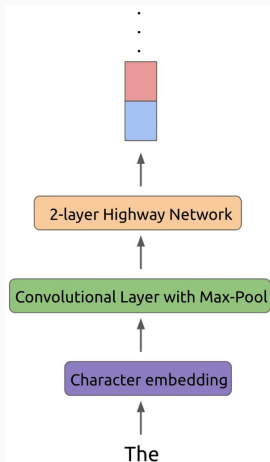
- Parando para pensar, o output de um modelo de língua em RNN/LSTM para uma certa palavra te dá um “embedding” para aquela palavra naquele contexto
 - Já que é feita uma operação matemática englobando tanto um embedding original como um vetor que codifica o contexto
 - Essa informação contextual é muito útil para captar variações semânticas de uma mesma palavra



Elmo - arquitetura

*“ELMo representations are **deep**, in the sense that they are a function of all of the internal layers of the biLM.”*

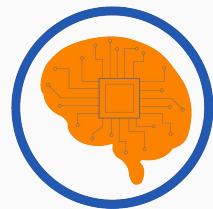
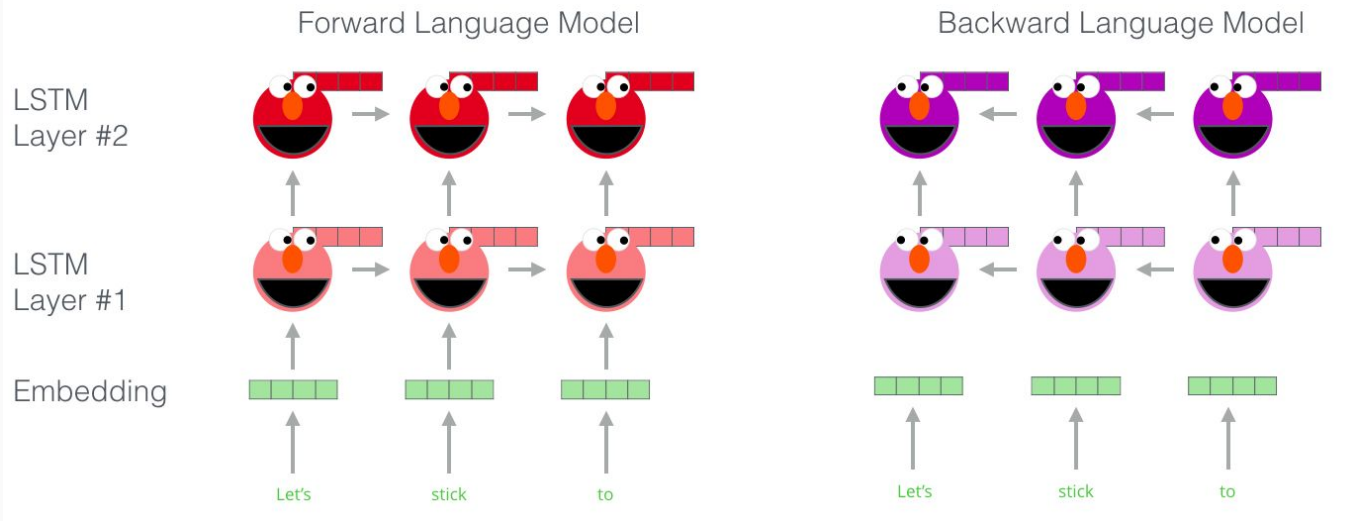
- Um modelo de língua bidirecional
 - Duas LSTMs de dois layers (um em cima do outro) independentes, uma prevendo a palavra seguinte dadas as palavras anteriores. A outra, prevendo a palavra anterior dadas as seguintes
- Embeddings “iniciais” um pouco mais complexos do que o usual



Elmo - arquitetura

*“ELMo representations are **deep**, in the sense that they are a function of all of the internal layers of the biLM.”*

Embedding of “stick” in “Let’s stick to” - Step #1



Elmo - o embedding final

"ELMo is a task specific combination of the intermediate layer representations in the biLM"

- Dependendo da tarefa, pode ser mais útil utilizar o output da LSTM de baixo ou de cima
 - Layer de baixo - mais "sintático" - POS, Parsing
 - Layer de cima - mais "semântico"

$$R_k = \{\mathbf{x}_k^{LM}, \vec{\mathbf{h}}_{k,j}^{LM}, \overleftarrow{\mathbf{h}}_{k,j}^{LM} \mid j = 1, \dots, L\}$$
$$= \{\mathbf{h}_{k,j}^{LM} \mid j = 0, \dots, L\},$$

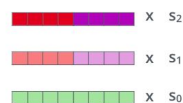
$$\text{ELMo}_k^{\text{task}} = E(R_k; \Theta^{\text{task}}) = \gamma^{\text{task}} \sum_{j=0}^L s_j^{\text{task}} \mathbf{h}_{k,j}^{LM}.$$

Embedding of "stick" in "Let's stick to" - Step #2

1- Concatenate hidden layers



2- Multiply each vector by a weight based on the task

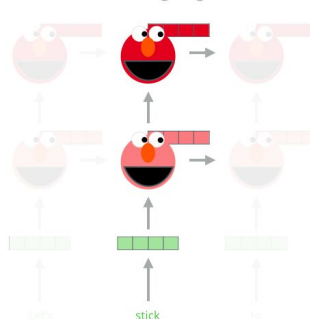


3- Sum the (now weighted) vectors



ELMo embedding of "stick" for this task in this context

Forward Language Model



Backward Language Model

