SCALIAN ACADEMY

Cultivons nos talents
Scalian Academy

DEEP LEARNING

Sequences et reseaux recurrents

LE TRAITEMENT DE SÉQUENCE

SCALIAN

- Représenter les mots et leur contexte comme vu précédemment aide, mais ne suffit pas à toutes les tâches

- NLP, Videos, Series temporelles, Audio, ADN, …. Sont autant de donnée pour lesquelles l'aspect séquentielle est important

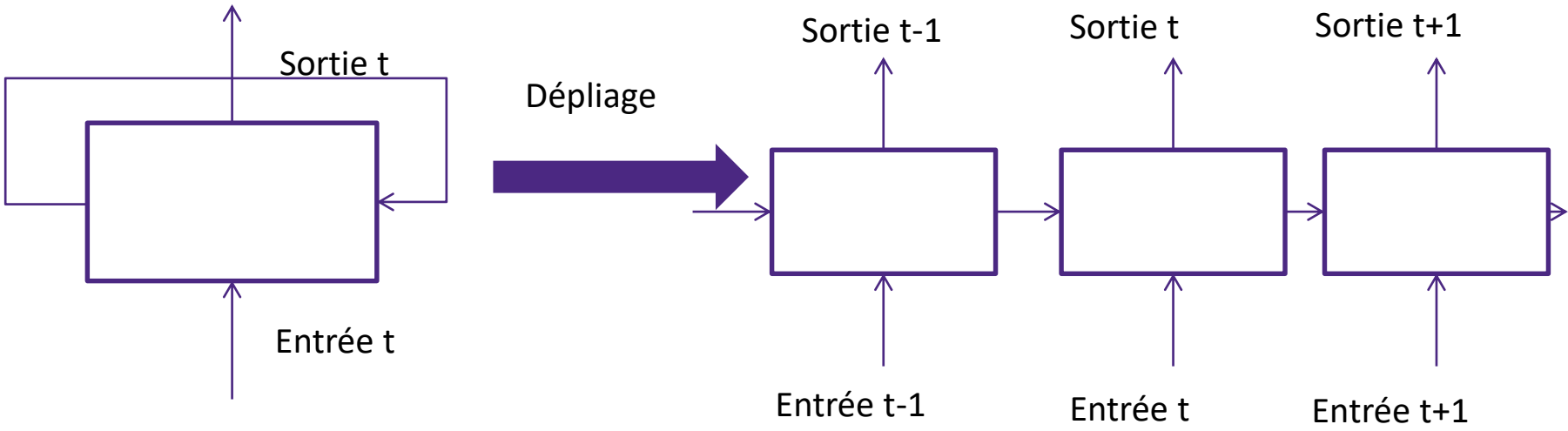
- Les premières approches étaient statistiques, et continuent de nos jours
  - Les réseaux de neurones n'ont pas autant remplacé l'état de l'art dans ce domaine qu'en traitement d'image
  - Des solutions et algorithmes variés : réseau bayésien, chaînes de Markoff, Dynamic time warping, SVM,…
  - Pour le NLP, principalement au moyen de réseaux de neurones

LES RÉSEAUX RÉCURRENTS

Deux types de récurrences, interne ou réseau ou externe, d'un step à un autre.

Peut prêter à confusion car souvent représentés pareils !

Pour clarifier, on représente souvent les deuxièmes avec du dépliage (« unrolling »)

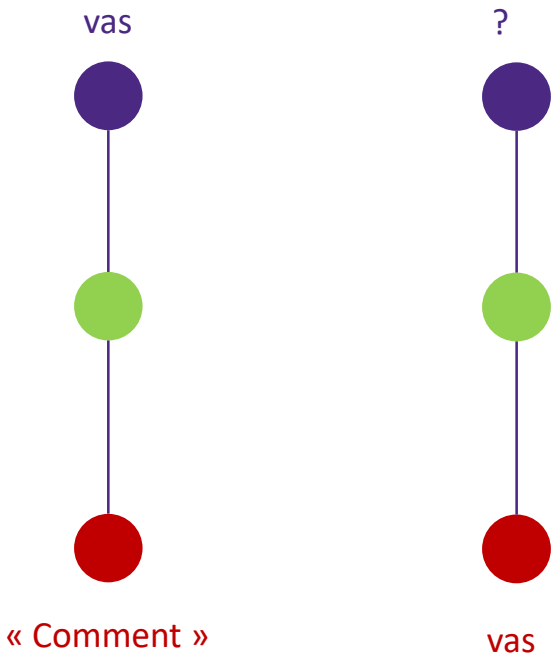


LE TRAITEMENT DE SÉQUENCE

EX : GENERATION DE SEQUENCE ET UNROLLING

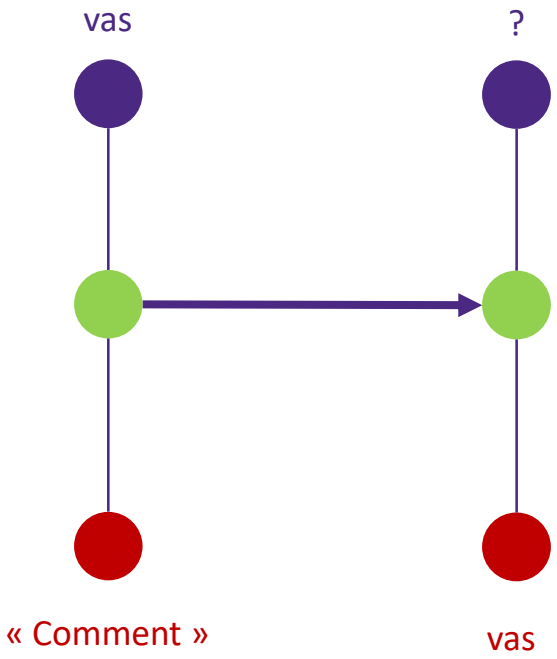
LE TRAITEMENT DE SÉQUENCE

EX : GENERATION DE SEQUENCE ET UNROLLING



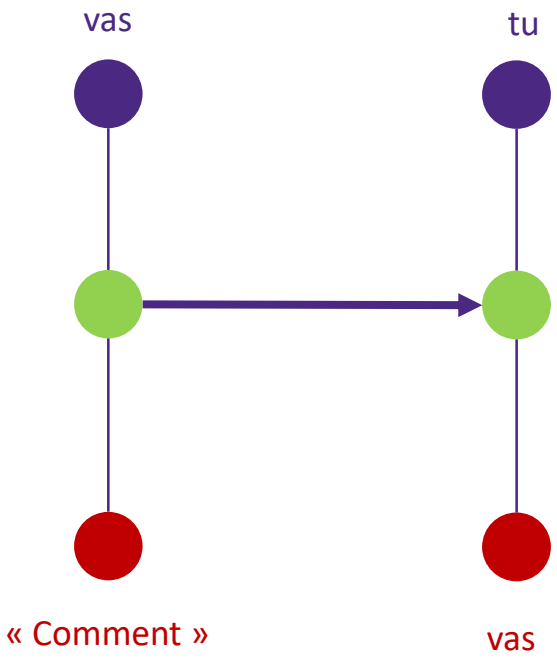
LE TRAITEMENT DE SÉQUENCE

EX : GENERATION DE SEQUENCE ET UNROLLING



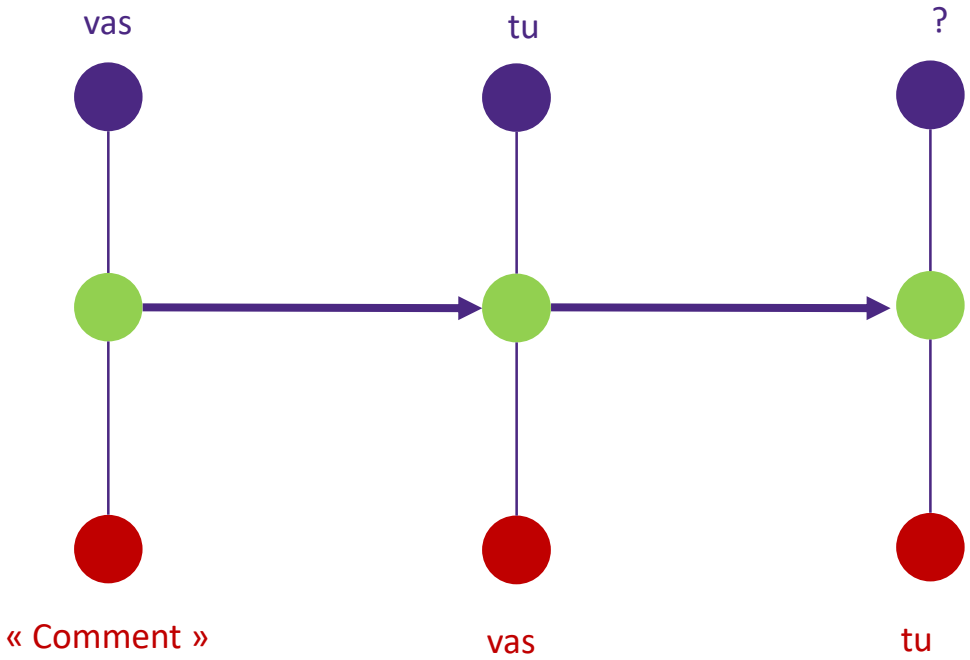
LE TRAITEMENT DE SÉQUENCE

EX : GENERATION DE SEQUENCE ET UNROLLING

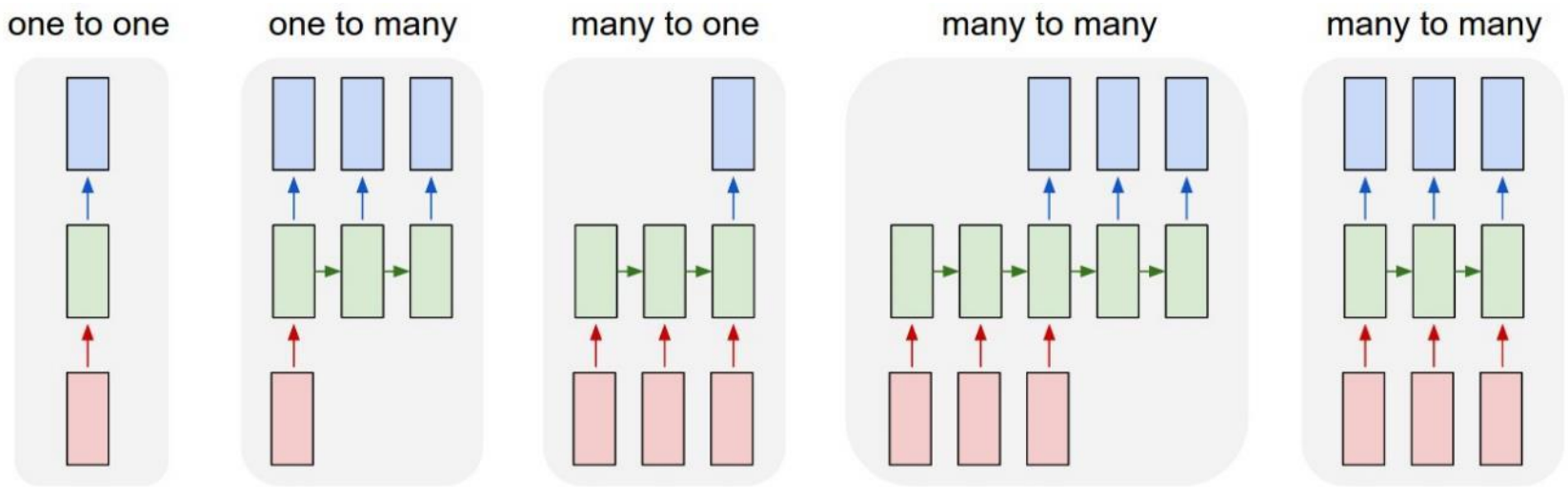


LE TRAITEMENT DE SÉQUENCE

EX : GENERATION DE SEQUENCE ET UNROLLING



LE TRAITEMENT DE SÉQUENCE

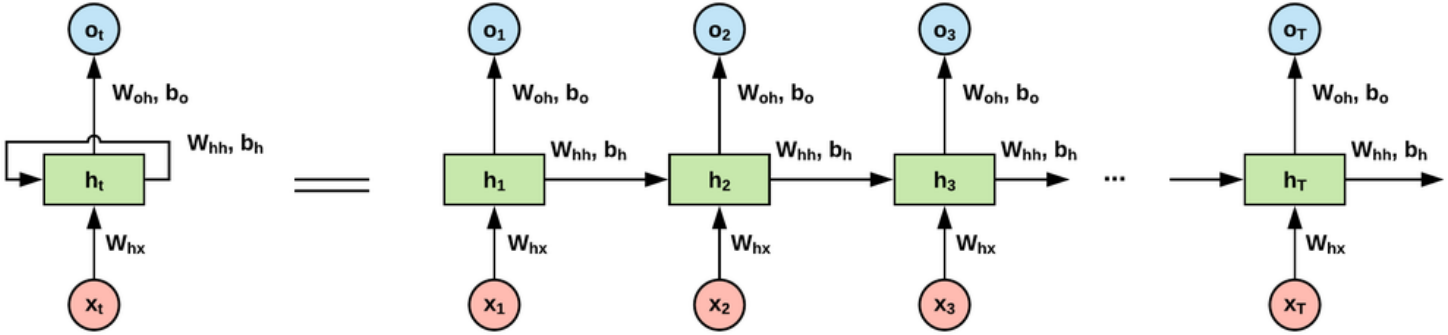


Question : dans l'exemple précédent, dans quel cas était-on ?

RECURENT NEURAL NETWORK

SCALIAN

COMMENT APPRENDRE DANS LE CAS OÙ UN EXEMPLE EST INFLUENCÉ PAR LES EXEMPLES PRÉCÉDENTS (SÉQUENCE) ?



Exemple 1 : J’ aime le ... Learning

Exemple 2 : [02, 10, 2015, 1500] [03, 10, 2015, 1205] [04, 10, 2015, 1820] ... [05, 10, 2015, 1900]

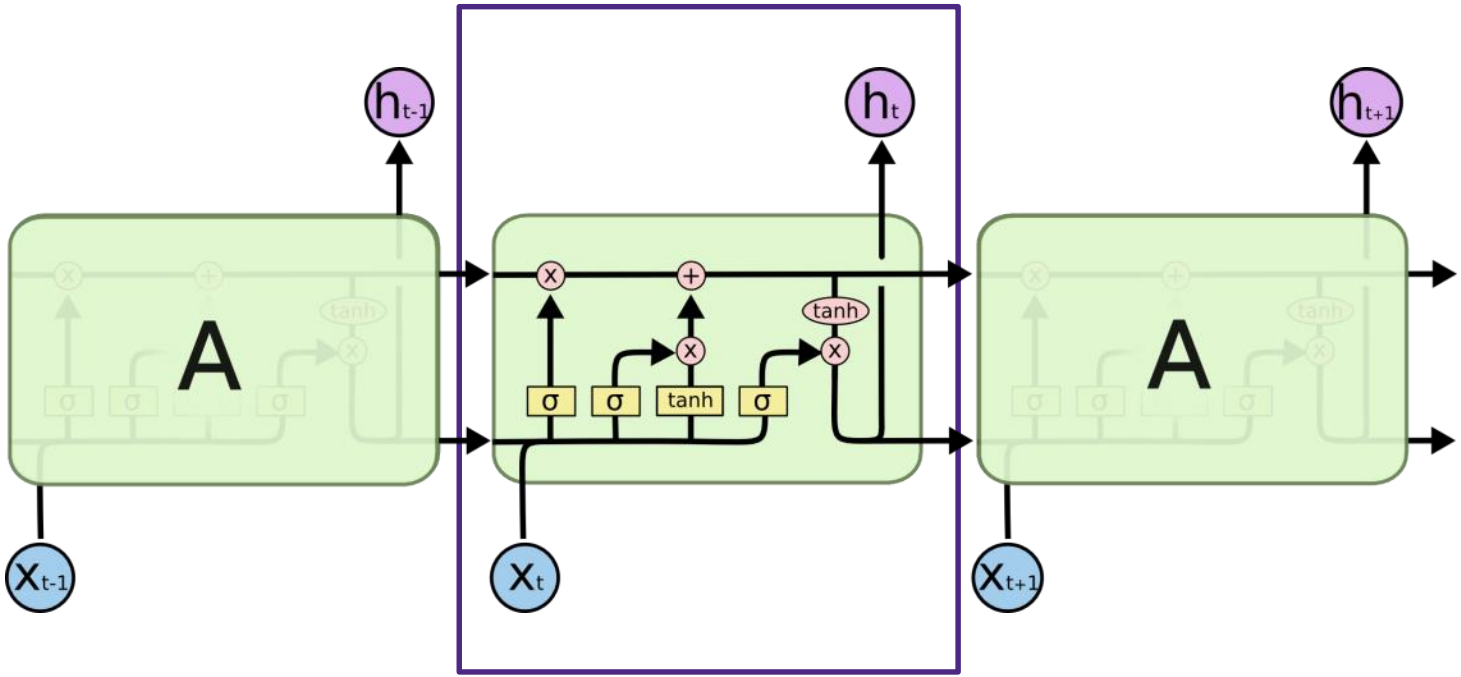
Jour, mois, année, valeur

LES LONG SHORT TERM MEMORY

SCALIAN

les LSTM, ou apprendre ce qu'il faut conserver des exemples précédents.

« Brique » LSTM



TANH, LA FONCTION DES RNN ?

SCALIAN

Relu rarement utilisé dans les RNN

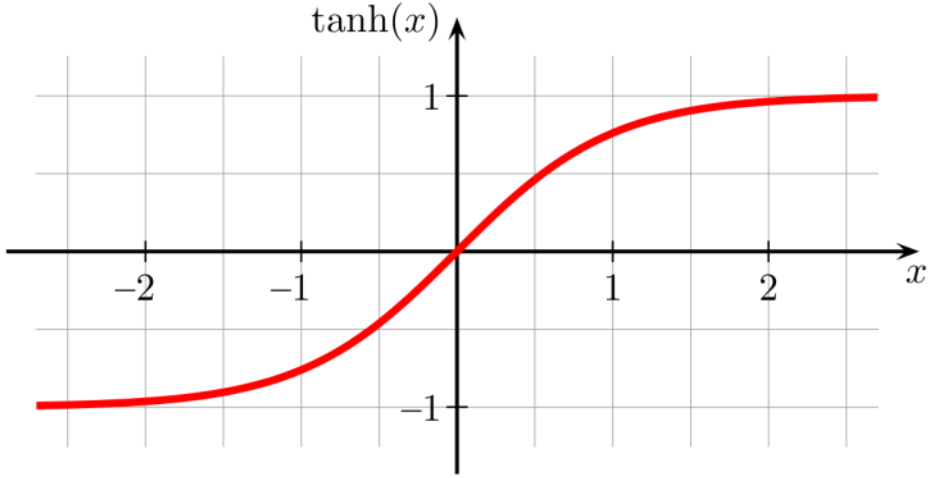
La tangente hyperbolique est très populaire

Avantages :

- **Calculs du gradient facile**
- **Gradient plus dur à disparaître**
- **Pour les RNNs, elle peut prendre des valeurs négatives, aide à stabiliser l'activité.**

Inconvénients :

- **Moins adaptée en sortie de réseau.**



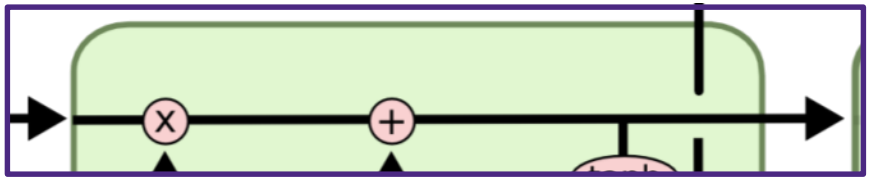
LES LSTM
LE CELL STATE

SCALIAN

Ligne du haut = « état » interne de la cellule :

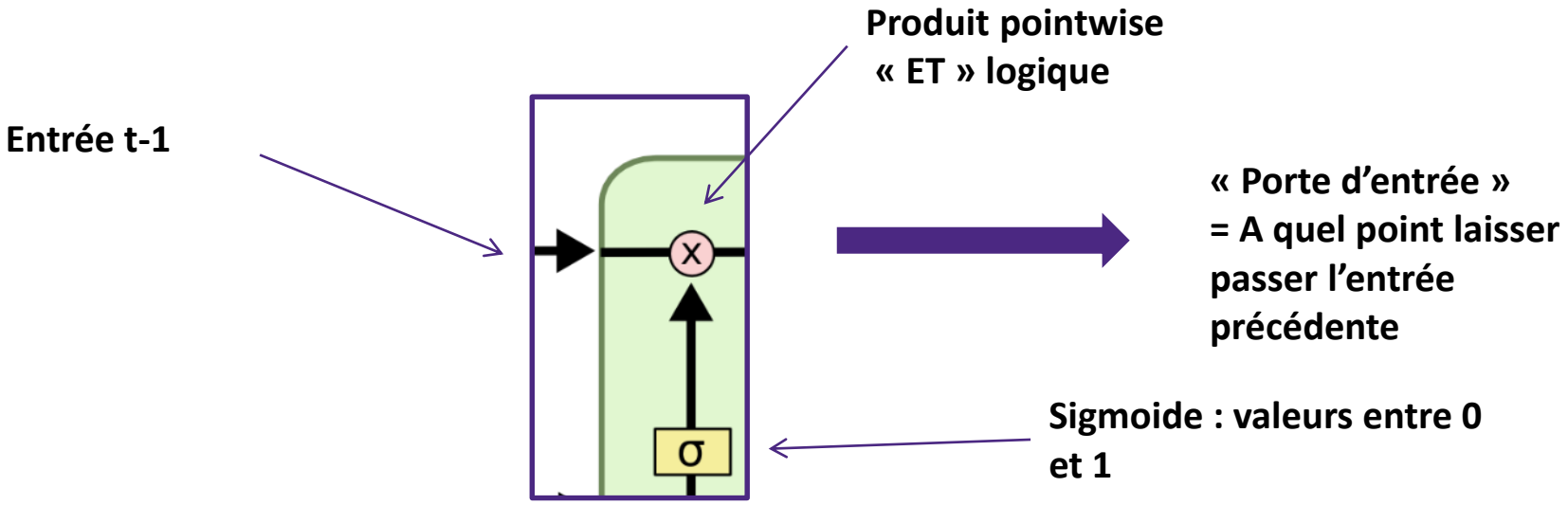
- **Peu d'intéraction avec le monde extérieur**
- **Difficilement déformé**
- **N'influence pas directement la sortie**
- **Principal flot d'information dans le temps**

« Cell State »



LES LONG SHORT TERM MEMORY

LES GATES



LES LONG SHORT TERM MEMORY

SCALIAN

3 portes dans notre LSTM !

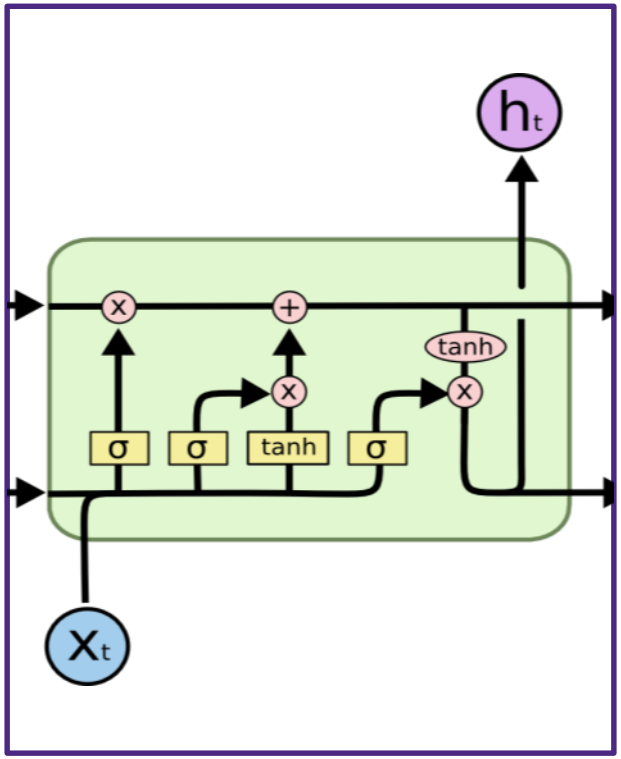
1. **Permet de contrôler quoi garder de notre état précédent**
   - **« porte d'oubli »**

2. **Permet de décider quelles informations sont à ajouter à l'état interne**
   - **« porte d'entrée »**

3. **Décide quelle partie de l'état prendre comme sortie**
   - **« porte de sortie »**

Intérêt de Tanh qui permet de soustraire !

« Brique » LSTM



DES VARIATIONS DE LSTM CELLS

SCALIAN

Beaucoup de variation de LSTM.

Principaux avantages :

- **pouvoir stocker l'information sur du long terme avec peu de changement**
- **Permettre à l'état interne d'influencer tout le temps le monde extérieur si nécessaire**
- **Peuvent être utilisé dans les deux sens pour apporter des informations sur les échantillons précédents et suivants sur du long terme !**

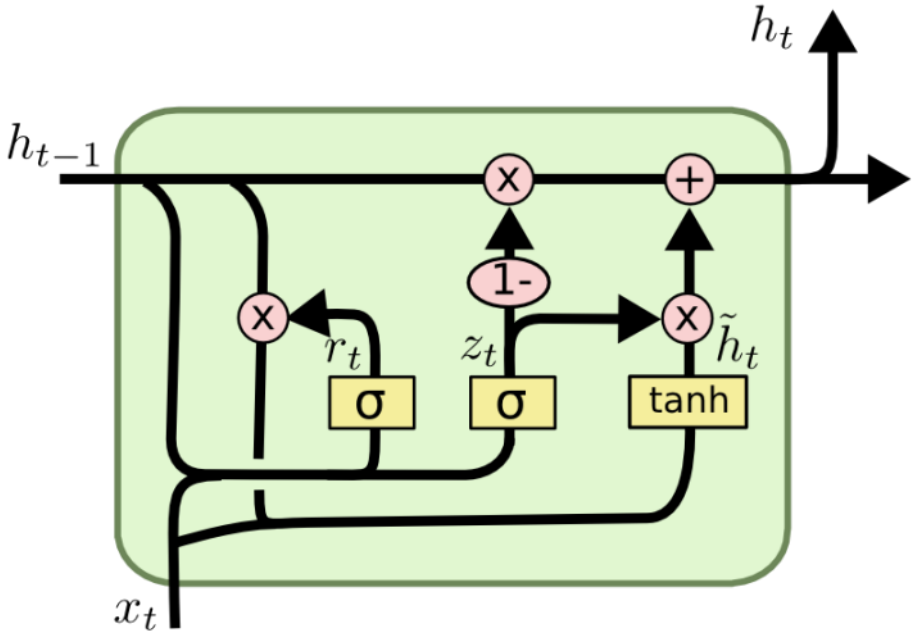
Variantes avec des « peephole » : laissent les gates « regarder » (être influencées) par l'état interne

Une variante notable des LSTMs: les Gated Recurrent Units (GRU)

LES GATED RECURRENT UNITS

GRU

- **Fusion de la porte d'oubli et d'entrée en une porte de « mise à jour »**
- **Fusionne l'état interne et la réponse de sortie**
- **Moins d'opérations et moins couteux**



En pratique, les LSTMs et les GRU ont quand même du mal à apprendre des dépendances à très long terme.

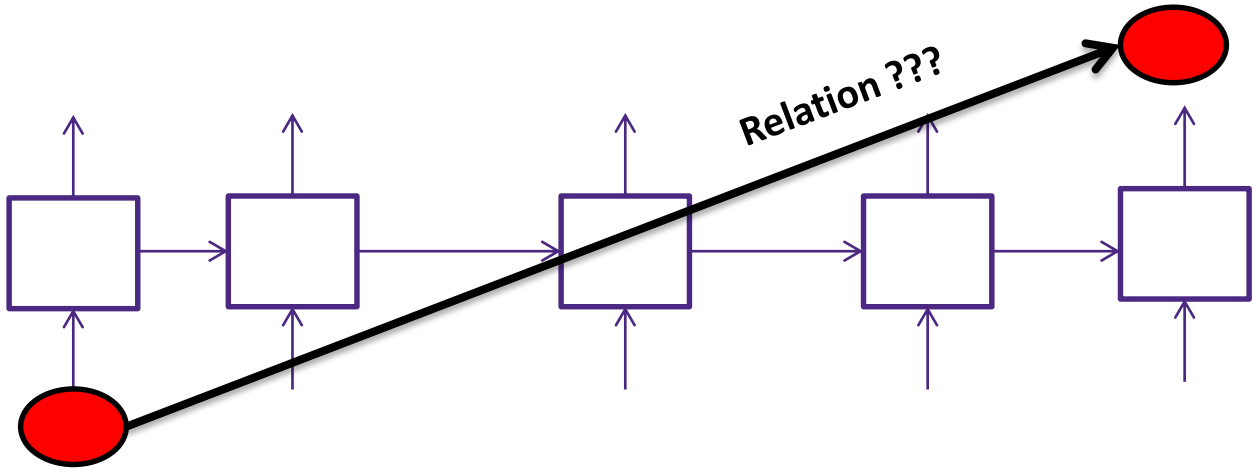
LES LIMITES DES RNN

LES LSTM

Problème des réseaux récurrents sur le très long terme: la disparition du gradient de l'erreur !

Long range similaire au deep learning, mais dans le temps !

« Preuve mathématique » : apprendre les variations de poids sur une tâche et conserver l'information pour apprendre dans le temps s'opposent (Bengio 1994)



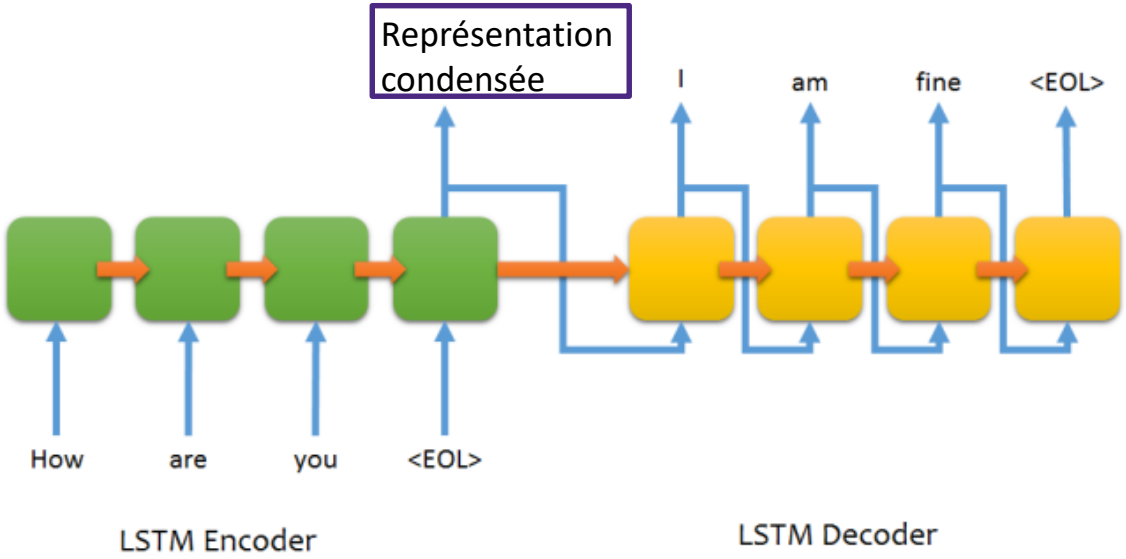
SEQ2SEQ

SCALIAN

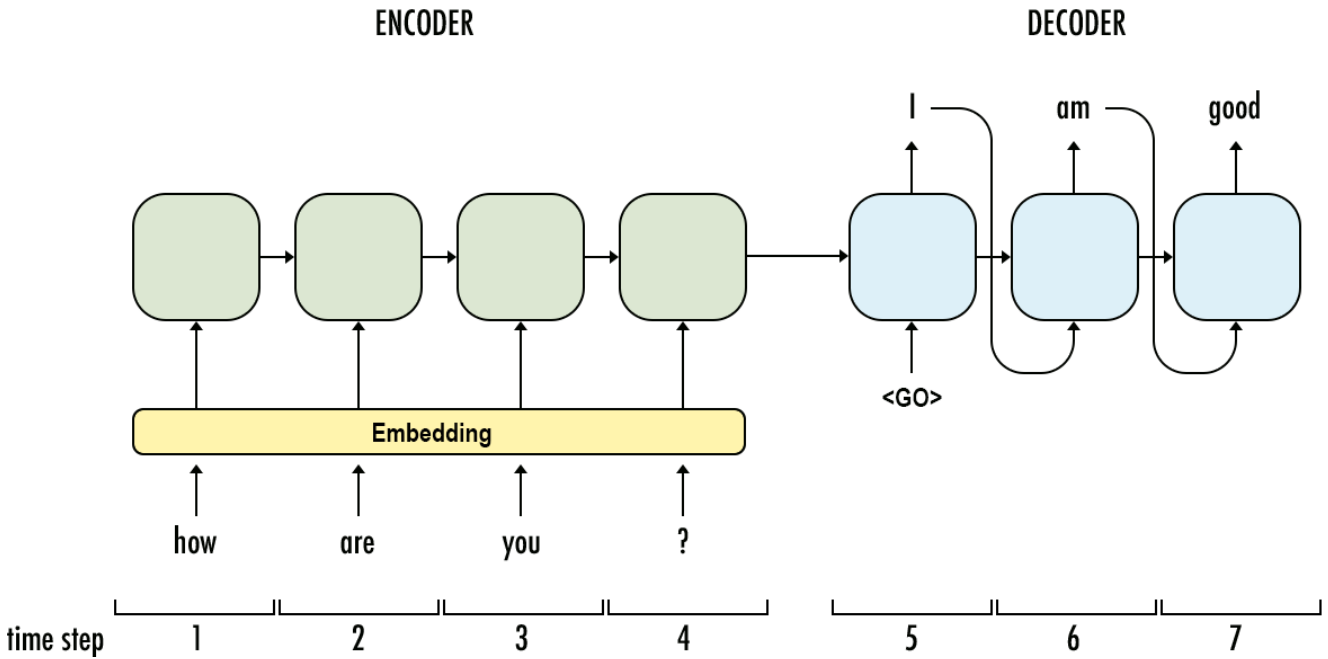
Solution Google, présente dans Tensorflow.

But : apprendre à associer une phrase de sortie à une phrase d'entrée sur le modèle

Applique le principe encoder/decoder aux lstm !



UN EXEMPLE : SEQ2SEQ

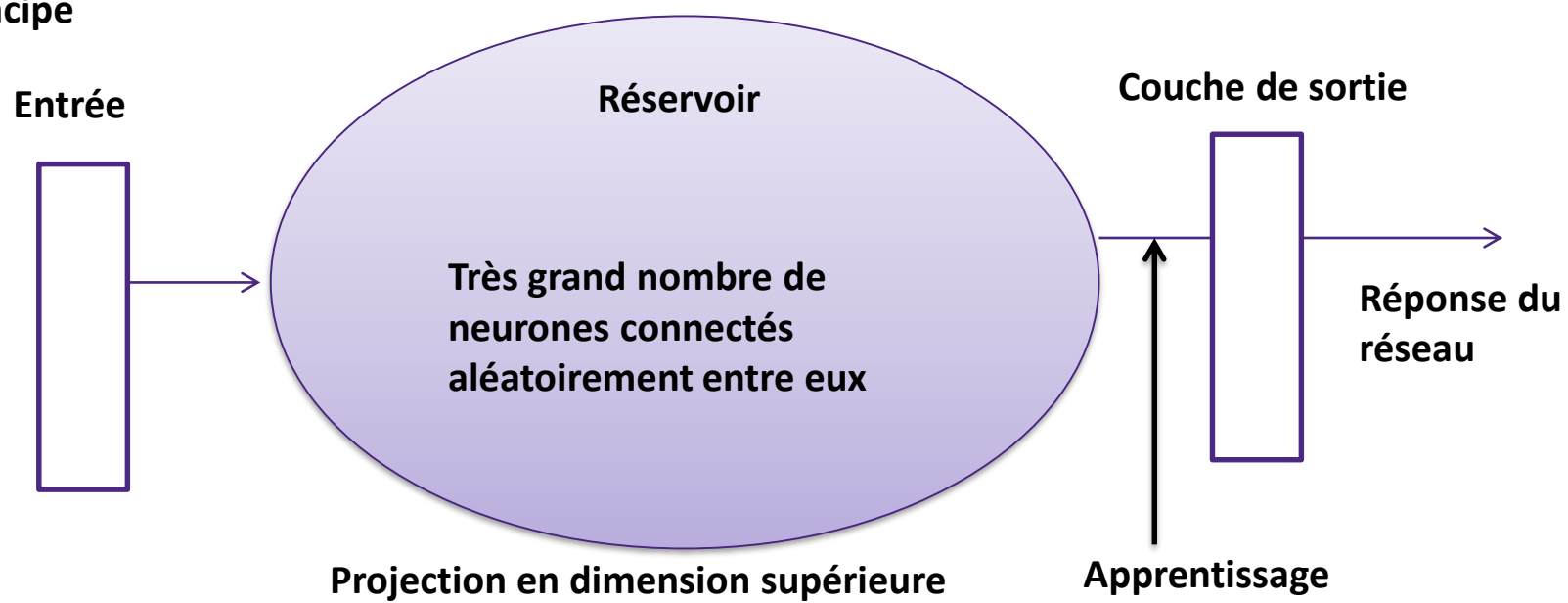


RESERVOIR COMPUTING

ECHO STATE NETWORK

Idée du reservoir computing : utiliser un système dynamique pour projeter l'entrée en grande dimension, rendant un apprentissage simple efficace.

Echo-state network : Réseau de neurone sparse (et donc peu couteux) basé sur ce principe



Apprentissage rapide car peu de neurones doivent apprendre.

Le réservoir permet de faire de l'apprentissage de séquence en gardant une image des derniers mots.

La matrice des poids aléatoire du réservoir ne doit pas être « trop remplie » pour converger

- **Doit respecter certains critères sur la valeur propre**
- **Il faut vivre dangereusement !**

Des variations comme le deep réservoir computing, liquid state machine, ou avec réservoir apprenant.

ENCODAGE SPARSE

SCALIAN

Avantages :

- **Beaucoup de modèles n'entraînent pas le réservoir et sont donc très rapides à entraîner.**
- **Naturellement adapté au traitement de séquence car la récurrence du réservoir conserve la trace des mots précédents.**

Inconvénient :

- **Beaucoup de neurones -> plus de calculs une fois en production.**
- **Nécessite de préserver certaines conditions sur la matrice de poids pour que les activités n'explosent pas (notamment sur sa « sparcité »).**
- **Très peu interprétable.**

CONCLUSION

SCALIAN

- **La prise en compte de séquence même à court terme à un sens en computer vision (dans le cas d'une vidéo)**

- **Certaines briques de bases des réseaux de neurones comme les lstm ou les gru permettent une représentation sur (un peu) plus long terme**

- **Comme souvent avec les réseaux de neurones, connaitre un petit ensemble de briques et de techniques permet de comprendre la grande majorité de ce qui se fait actuellement.**