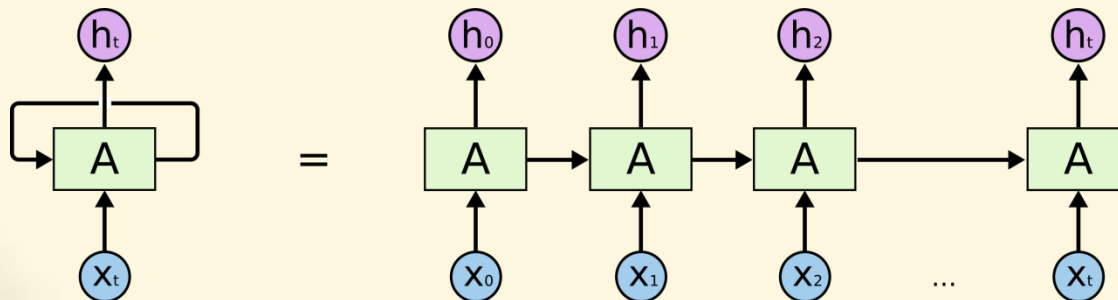


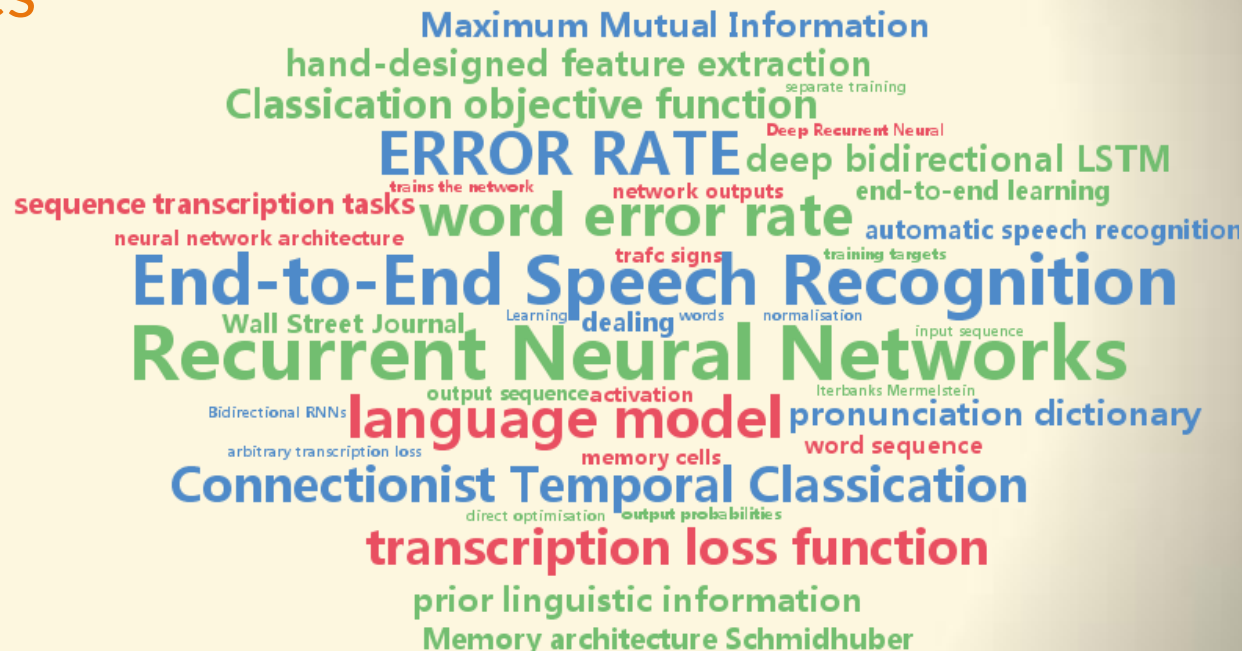
Parcours Data Scientist

Projet 8 : Veille Technologique Recurrent Neural Networks



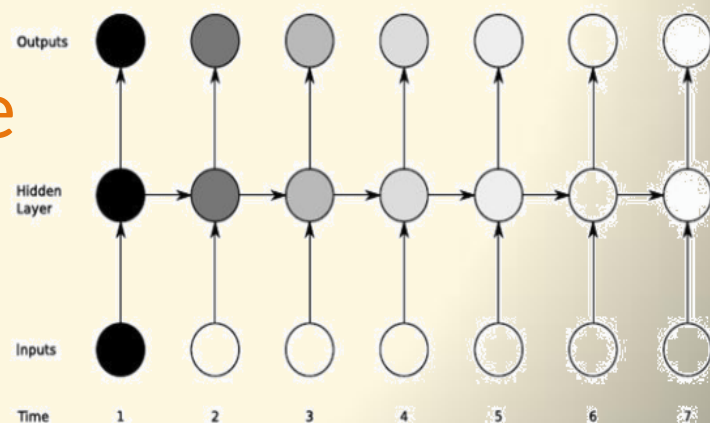
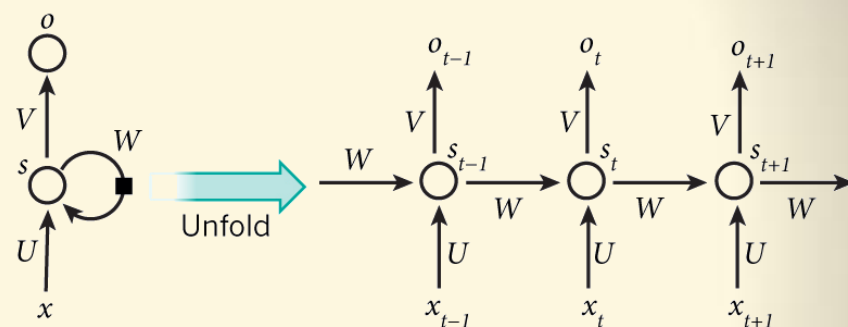
Sommaire

- Principe du RNN
 - Problèmes liés
 - Utilisation
- Etat de l'art
 - Simple RNN
 - LSTM
 - GRU
 - QRNN
- Evaluation
- Conclusion



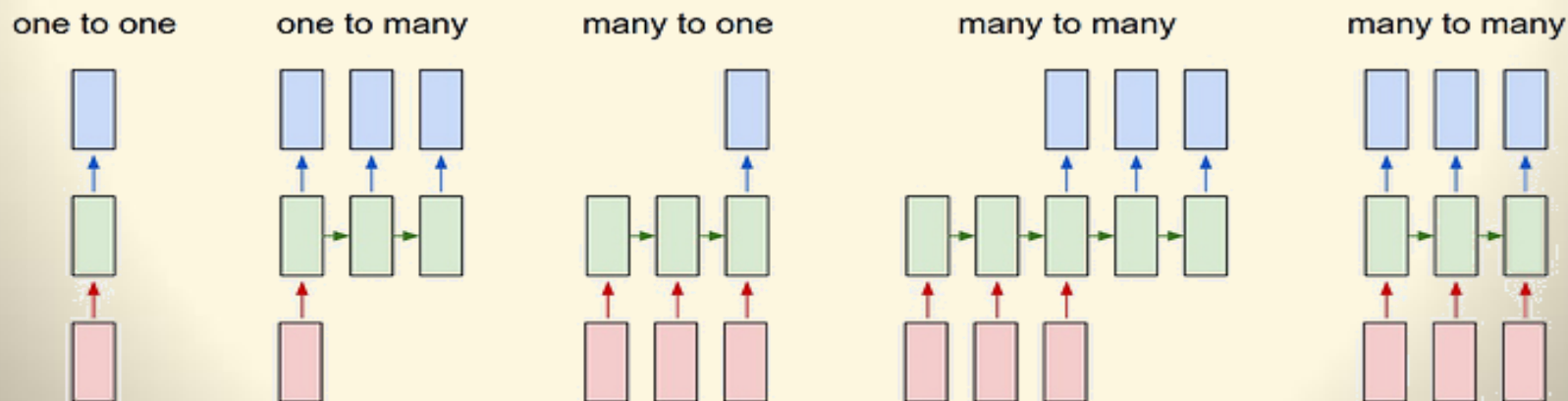
Principe du RNN

- Prédiction sur des données temporelles
 - Sorties récurrentes
- Utilisation multiples
 - Retards
 - analyse de texte
 - Traduction
 - ...
- Back propagation différentielle
 - BP dans le temps et l'espace
- Problème
 - Vanishing Gradient



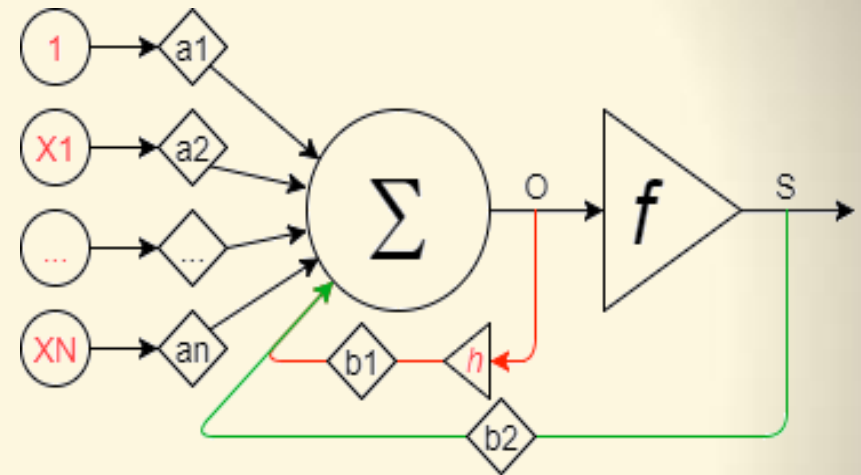
Principe du RNN

- Utilisation
 - 1-to-1 : Classification simple
 - 1-to-N : Générateur
 - N-to-1 : Classification « complexe »
 - N-to-N : Traduction / multi Classification



Etat de l'art

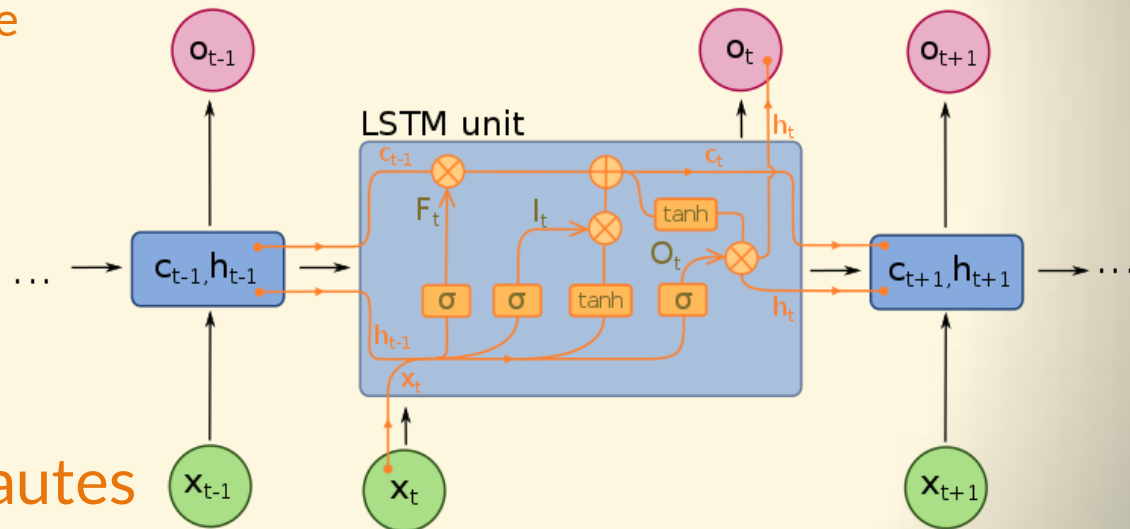
- Simple RNN
 - 1970's
 - Pb Vanishing Gradient
 - Rapide (calcul simple)
 - Peu utilisé



- $S(t) = f\left(\sum_{inputs} a_i * x_i(t) + h\left(\sum_{inputs} a_i * x_i(t-1)\right)\right)$
- $S(t) = f\left(\sum_{inputs} a_i * x_i + b2 * (S(t-1))\right)$

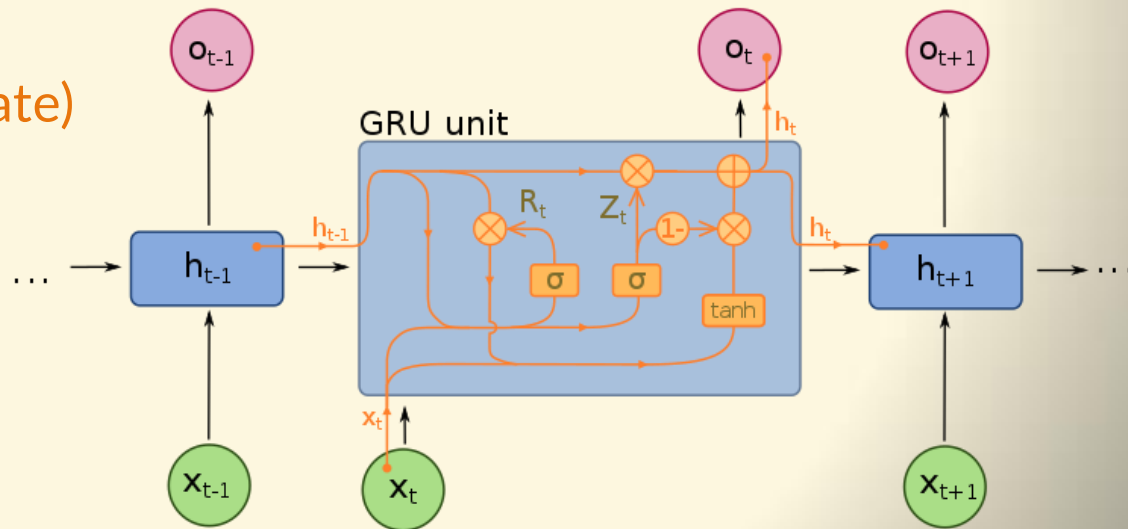
Etat de l'art

- LSTM 1997
 - 2 portes + 2 Etats
 - Hidden & Cell State
 - Input & Output Gate
- LSTM 2000
 - 3ème porte
 - Forget Gate
- Performances très hautes
- Gourmant en calcul
- Nombreuses variations



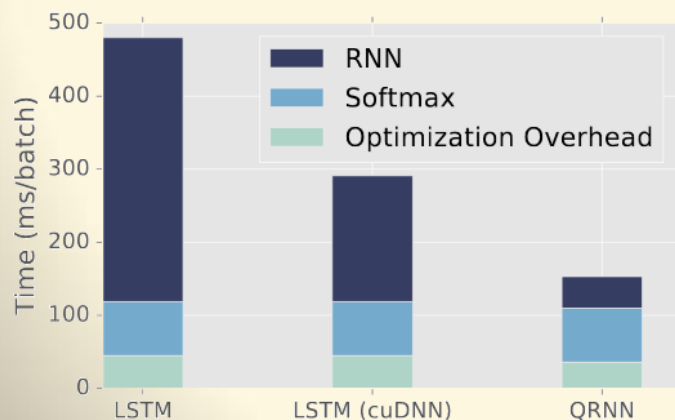
Etat de l'art

- GRU
 - 2014
 - Plus rapide que LSTM
 - Même performances
 - Basé sur LSTM
 - 1 état (Hidden State)
 - 2 portes
 - Update Gate
 - Reset Gate



Etat de l'art

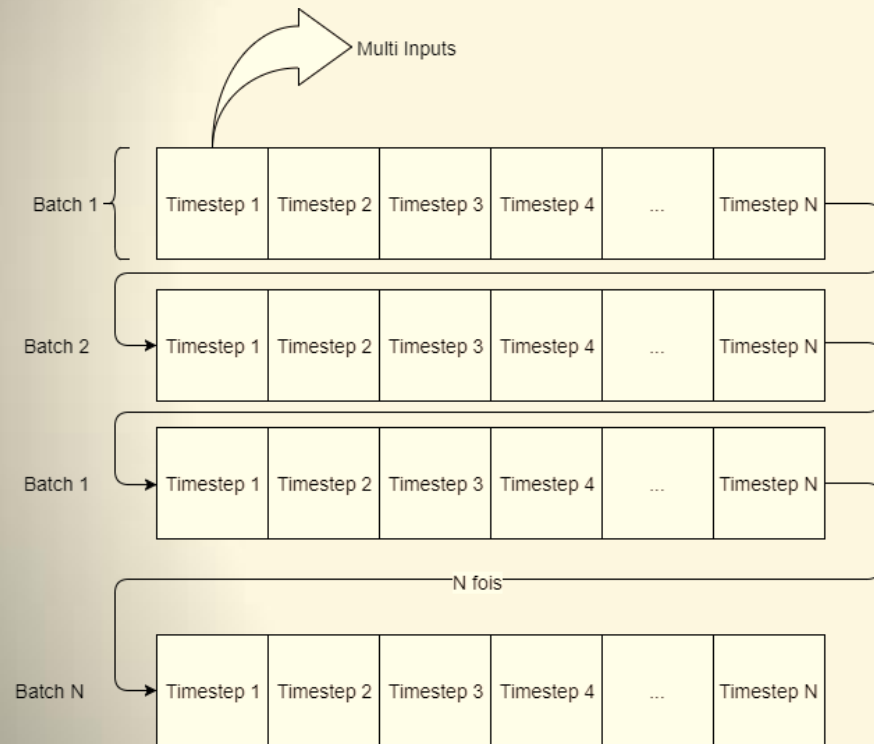
- QRNN
 - 2016/2017 (<https://arxiv.org/abs/1611.01576>)
 - Nouvelle architecture (Parallélisme des calculs)
 - Plus rapide que LSTM/GRU
 - Performances encore basses



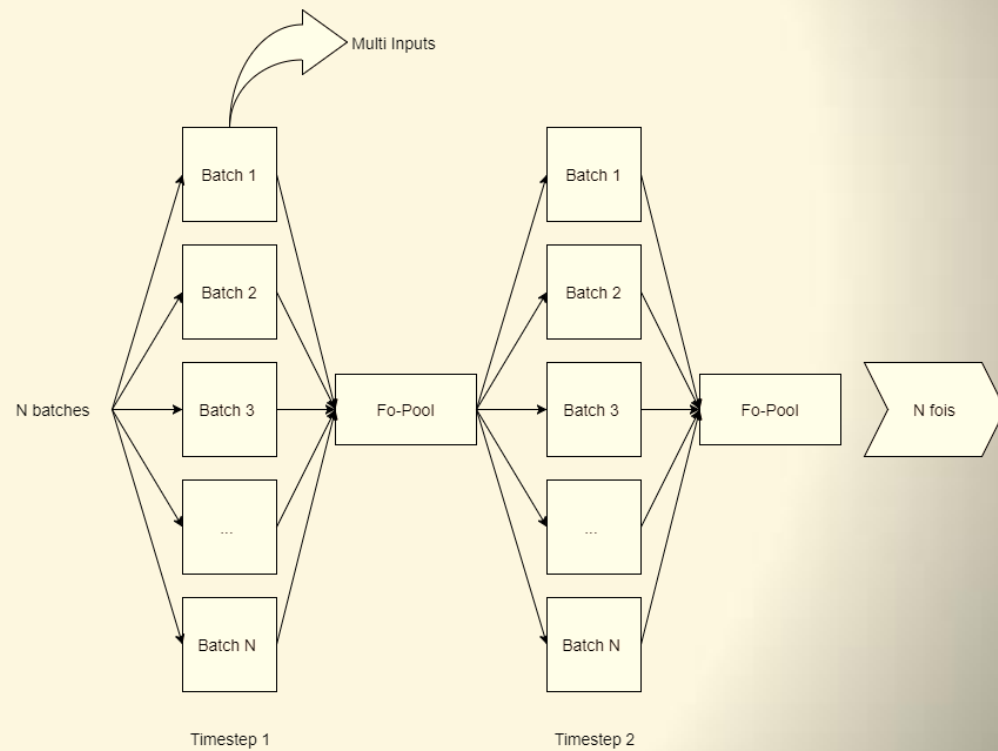
		Sequence length				
		32	64	128	256	512
Batch size	8	5.5x	8.8x	11.0x	12.4x	16.9x
	16	5.5x	6.7x	7.8x	8.3x	10.8x
	32	4.2x	4.5x	4.9x	4.9x	6.4x
	64	3.0x	3.0x	3.0x	3.0x	3.7x
	128	2.1x	1.9x	2.0x	2.0x	2.4x
	256	1.4x	1.4x	1.3x	1.3x	1.3x

Etat de l'art

- QRNN



SRNN/LSTM/GRU



QRNN

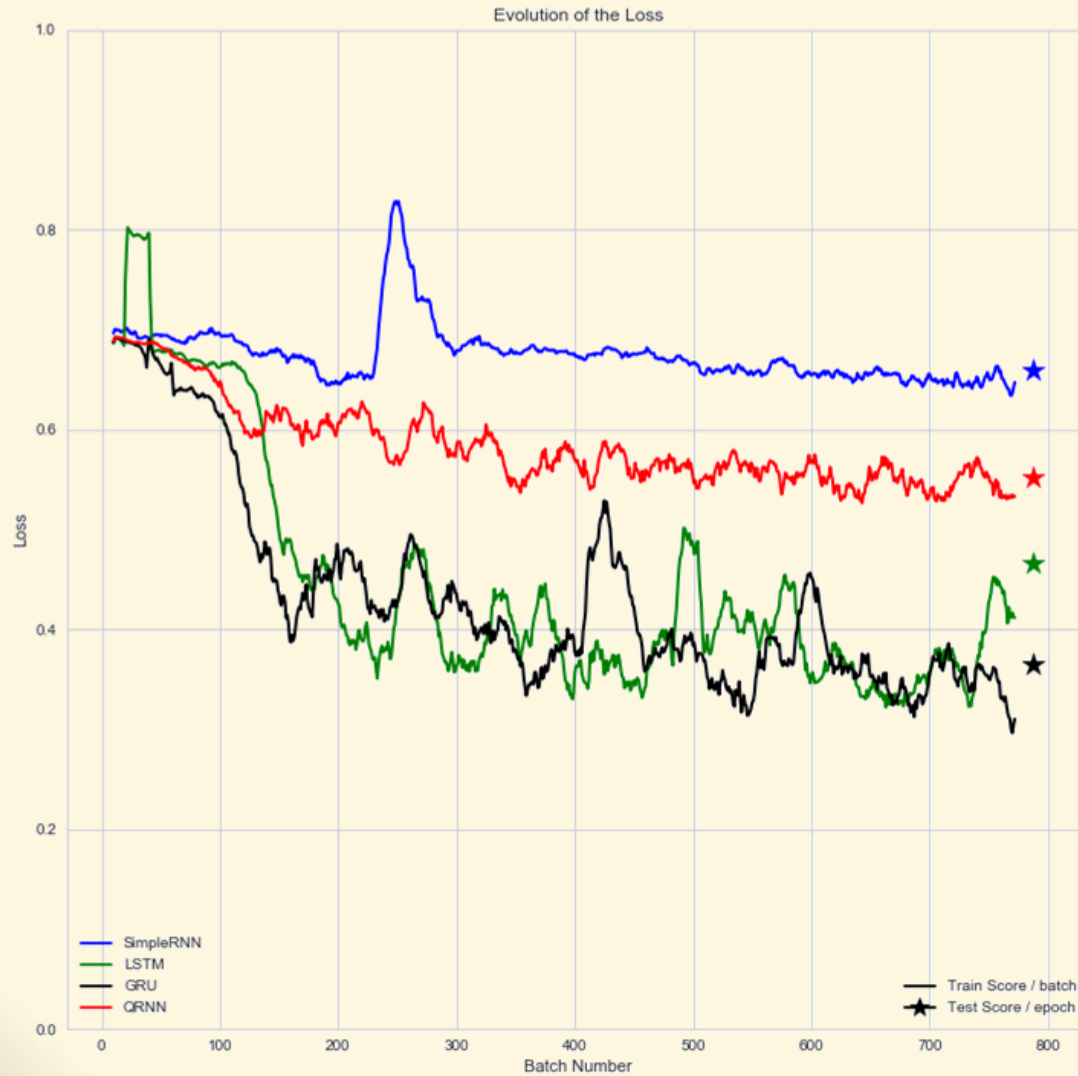
Evaluation

- Dataset : Large Movie Review Dataset
- Type : Analyse de Sentiments
- Contenu : 25k commentaires train + 25k test
 - (50 % positifs et Négatifs)
 - Longueur de la séquence à choisir
 - Dataset préparé
- Evaluation à topologie identique
 - 1 Embedding Layer (1 -> 128 dimensions)
 - 1 Cellule SRNN/LSTM/GRU/QRNN
 - 1 Layer FC + Sigmoid
 - Optimiser : Adam
 - Métriques : Accuracy, Loss, Temps
 - 1 Epoch (Overfitting)

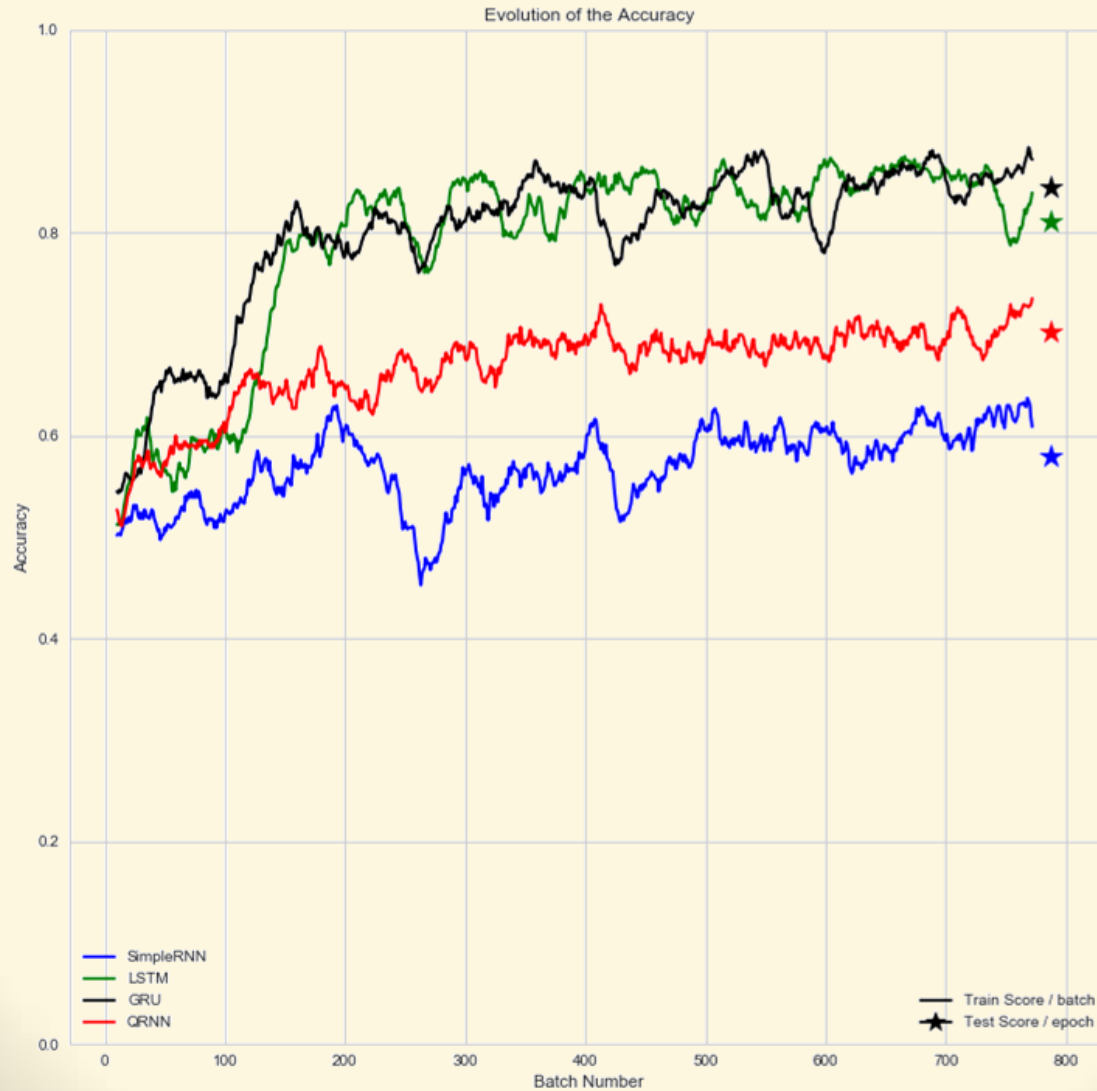
Evaluation



Evaluation



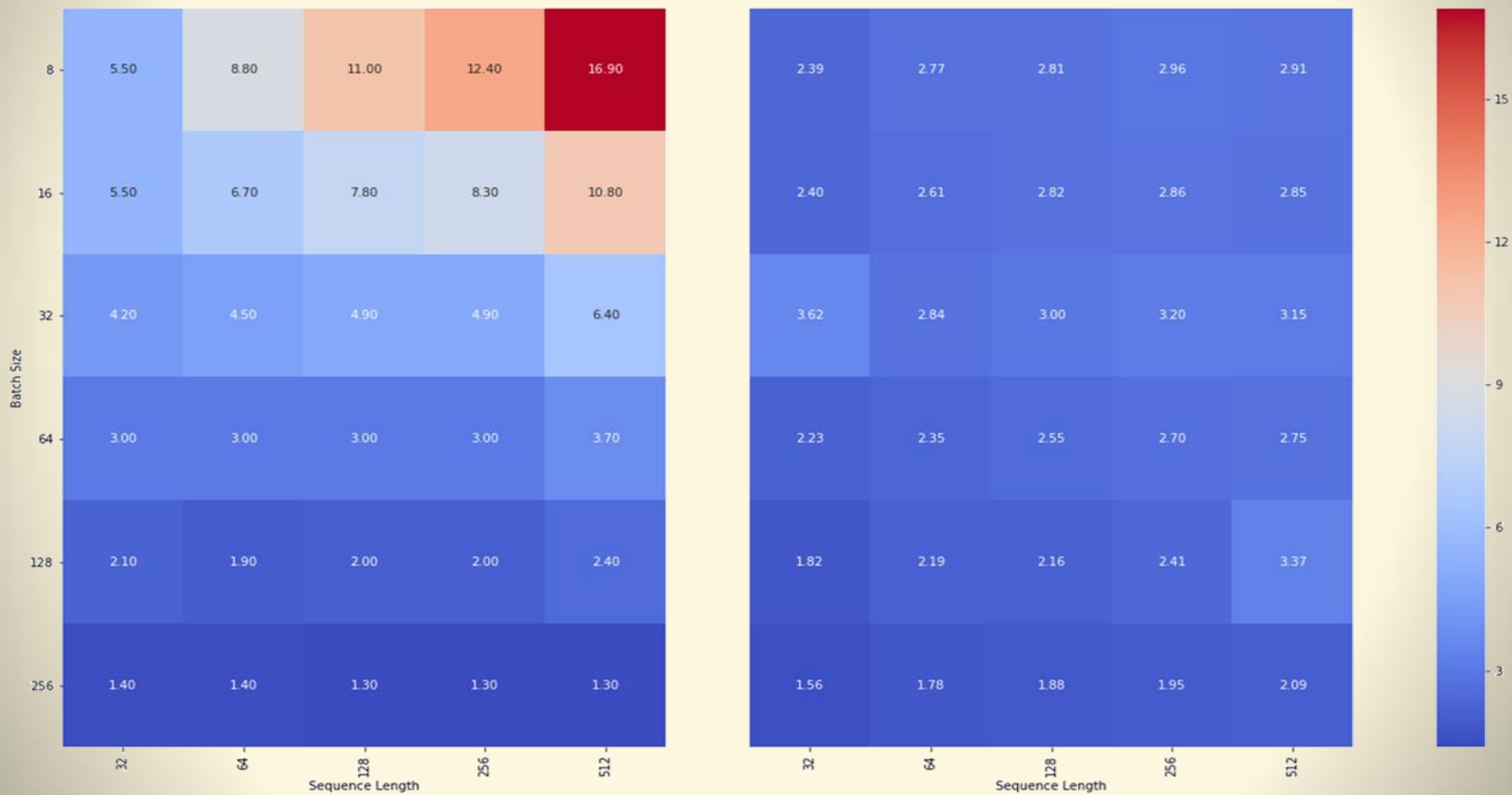
Evaluation



Evaluation

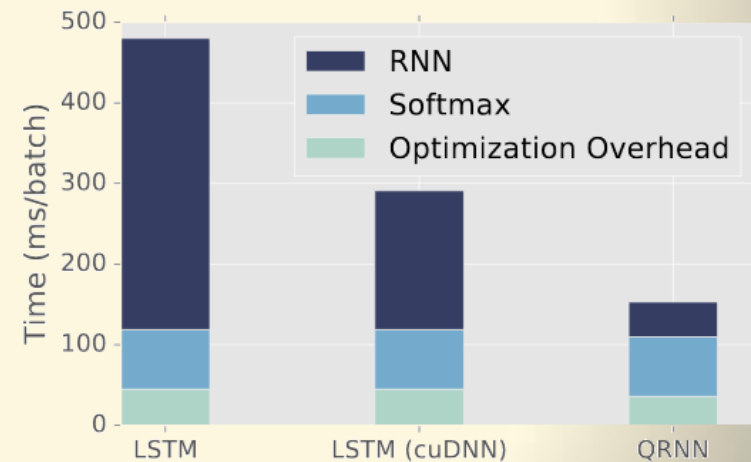


Evaluation



Evaluation

- Pas les Performances souhaitées
 - Possibilité d'optimisations importantes
 - Régularisations
 - Stride/Kernel
 - Dropout (1D)
 - Gradient Clipping
 - Modèles récent
 - Evolutions possible (LSTM)
- Pas le gain de temps évalué
 - Différents critères
 - Différents GPU ?
 - Modèle non optimisé (Github) ?



Conclusion

- Forte évolutions en 50 ans
- Champ d'application très large
- Très démocratisé (traduction, Analyse de sentiments, classification)
- 30 ans de flottements
- Encore limite (musique : GANs)
- LSTM très majoritaire
- QRNN encore un peu jeune mais prometteur
- Le sythetic gradient : novateur contre BPTT ?

