

RÉSEAUX DE NEURONES RÉCURRENTS

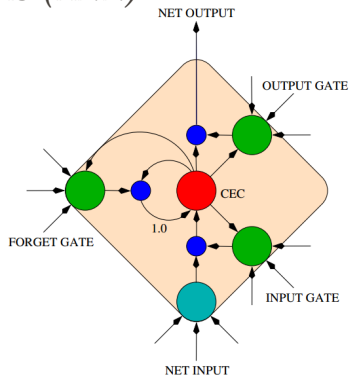
RECURRENT NEURAL NETWORKS (RNN)

Projet de Traitement de l'Écrit

20 février 2015

G4E Thomas ROBERT

Université de Rouen



Sommaire

1 ARCHITECTURES CLASSIQUES DE RNN

- Historique
- Architectures
- Apprentissage
- Applications

2 NOUVELLES ARCHITECTURES DE RNN

- Historique
- Principe
- Architectures
- Apprentissage
- Applications

3 CONCLUSION

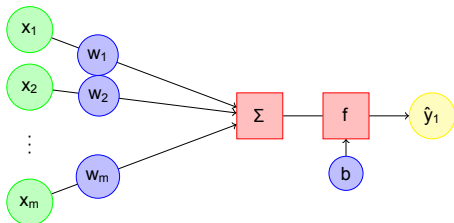
4 BIBLIOGRAPHIE

ARCHITECTURES CLASSIQUES DE RNN

- **1943** : Le neurone formel [**McCulloch43**]
Une représentation mathématique d'un neurone biologique.
- **1949** : Loi d'adaptation [**Hebb49**]
- **1958** : Le perceptron [**Rosenblatt58**]
- **1986** : Le réseau récurrent de Jordan et Elman [**Jordan86**]

Historique

Le neurone formel



- x_i : entrées
- w_i : poids
- f : fonction d'activation
- b : biais
- \hat{y} : sortie du neurone

FIGURE 1: Le neurone formel – Source :

Cours R. HÉRAULT & P. LERAY

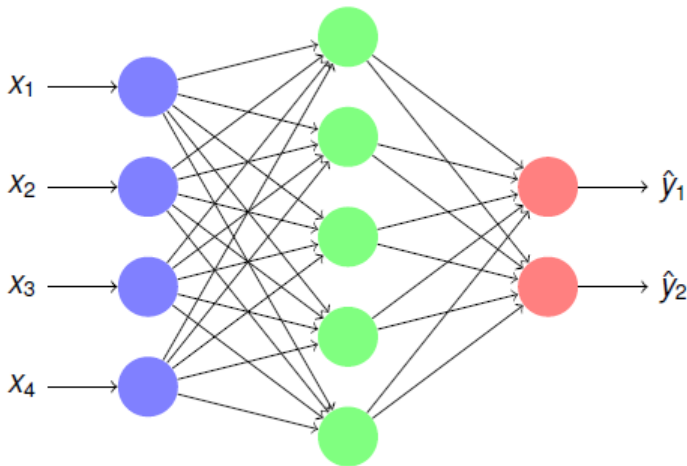


FIGURE 2: Réseau de neurones feed-forward – Source : Cours R. HÉRAULT & P. LERAY

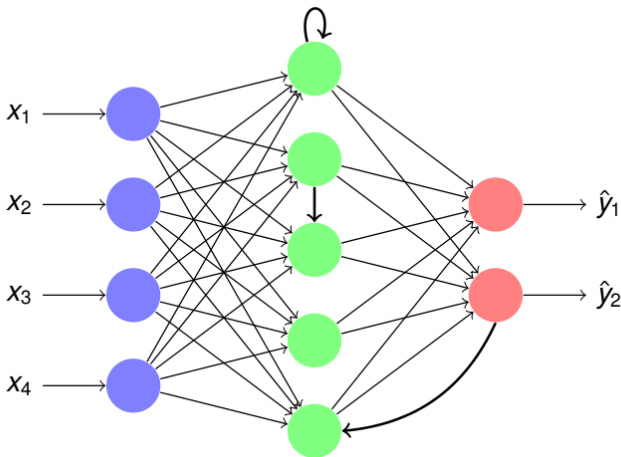
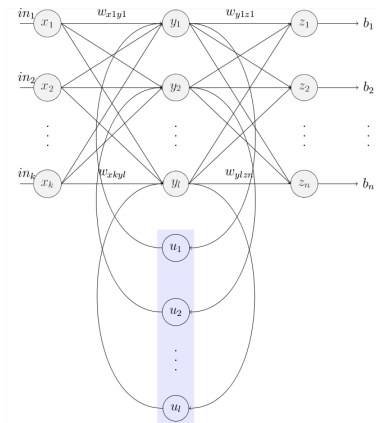


FIGURE 3: Réseau de neurones récurrent – Source : Cours R. HÉRAULT & P. LERAY

FIGURE 4: RNN d'Elman – Source : [wikimedia.org](https://commons.wikimedia.org/wiki/File:Elman_RNN_diagram.png)

- La couche $u_1 \dots u_n$ sert de mémoire pour l'état précédent

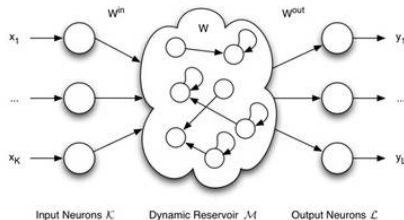


FIGURE 5: Echo State Network – Source : Université de Ulm

- Par JAEGER en 2001 [Jaeger01]
- Pas d'architecture précise
- Réservoir de neurones faiblement connectés aléatoirement entre eux (*Reservoir computing*)

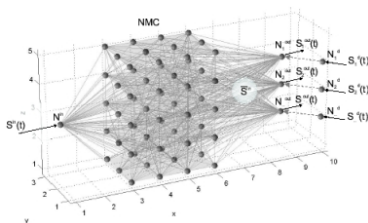


FIGURE 6: Liquid State Machine – Source : F. PONULAK

- Par MAASS, NATSCHLÄGER et MARKRAM en 2002 [Maass02]
- Principe proche des *Echo State Network*
- Réseau de neurones à décharge (*spiking neural network*)
- Pas d'architecture précise
- Réservoir de neurones connectés aléatoirement entre eux

- Les deux méthodes les plus communes
 - Real Time Recurrent Learning (RTRL) [**Robinson87**]
 - BackPropagation Through Time (BPTT) [**Williams89**]

- On calcul le gradient de l'erreur par rapport aux poids et on met à jour les poids à chaque pas de temps en conséquence
- Complexité $\mathcal{O}((n + l)^4)$ (n entrées, l sorties)

- Déplier le réseau pour l'approximer par un réseau non récurrent
- On applique une backpropagation classique sur le réseau déplié

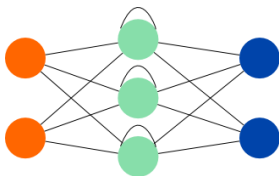


FIGURE 7: Exemple de RNN à l'instant t

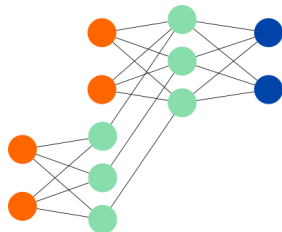


FIGURE 8: RNN déplié jusqu'à l'instant $t-1$

- Défaut : Difficile d'utiliser cette méthode sur toute la période
 - On limite aux k dernières itérations : on oublie des informations
- Complexité $\mathcal{O}(t \times n^2)$ (n neurones, t étapes temporelles)

- Très peu d'applications publiées basées sur des RNN simples
 - Apprentissage de trajectoires d'espace d'état dans un RNN [Pearlmutter89]
 - Résolution de l'équation de Sylvester à coefficients variables dans le temps [Zhang02]
 - Modélisation prosodique du Mandarin et application à la reconnaissance de la parole [Wang02]

NOUVELLES ARCHITECTURES DE RNN

- Les réseaux de neurones récurrents classiques conservent l'information sur une courte durée (une dizaine d'itérations maximum)
- **1991** : Démonstration de ce problème par la thèse de Sepp Hochreiter (encadré par Jürgen Schmidhuber) [**Hochreiter91**]
- Recherche de solutions à ce problème par Schmidhuber & Hochreiter

- **1997** : HOCHREITER et SCHMIDHUBER publient un article nommé “*Long short-term memory*” [Hochreiter97] (fig. 9)
- **2000** : Ajout de la « forget gate » au modèle par GERS, SCHMIDHUBER et CUMMINS [Gers00] (fig. 10)

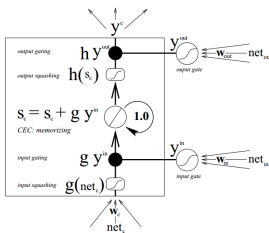


FIGURE 9: LSTM original

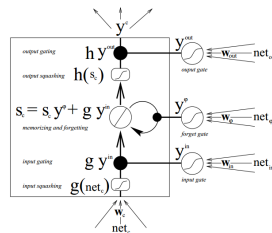


FIGURE 10: Ajout de la « forget gate »

Le travail sur les LSTM a été poursuivi par Schmidhuber et ses thésards et collaborateurs, en particulier Alex Graves.

- **2001** : Apprentissage des langues [**Gers01**]
- **2003** : Apprentissage de timings [**Gers03**]
- **2005** : Reconnaissance de phonèmes [**Graves05a** ; **Graves05b**]
- **2006** : Apprentissage de données non-segmentées [**Graves06**]
- **2009** : Reconnaissance de l'écriture [**Graves09a** ; **Graves09b**]
- **2013** : Reconnaissance de la parole [**Graves13a**]

Livre de référence : GRAVES, *Supervised sequence labelling with recurrent neural networks* [**Graves12**]

Principe

Cellule d'un LSTM

- Cellule mémoire au centre
- Portes contrôlant le flux d'information entrant / sortant / restant dans la cellule.
- Permettent de conserver de l'information durablement si besoin.
- Peuvent être vues comme des opérations :
 - Ecriture (*input gate*)
 - Lecture (*output gate*)
 - Réinitialisation (*forget gate*)

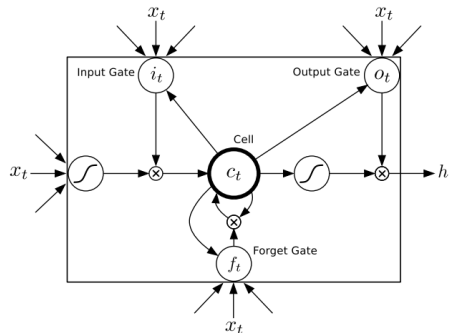


FIGURE 11: Cellule LSTM – Source :
[Graves13b]

Exemple de comportement possible

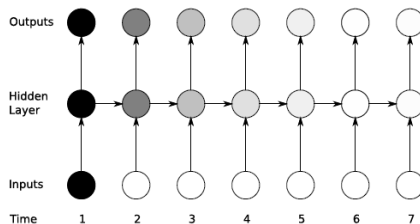


FIGURE 12: Exemple pour un RNN – Source : [Graves12]

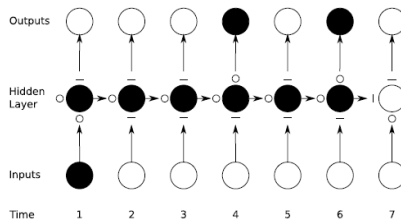


FIGURE 13: Exemple pour un LSTM – Source : [Graves12]

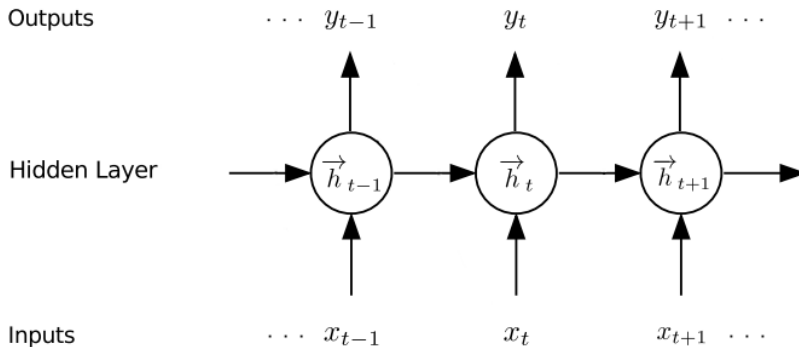


FIGURE 14: RNN classique – Source : [Graves13b]

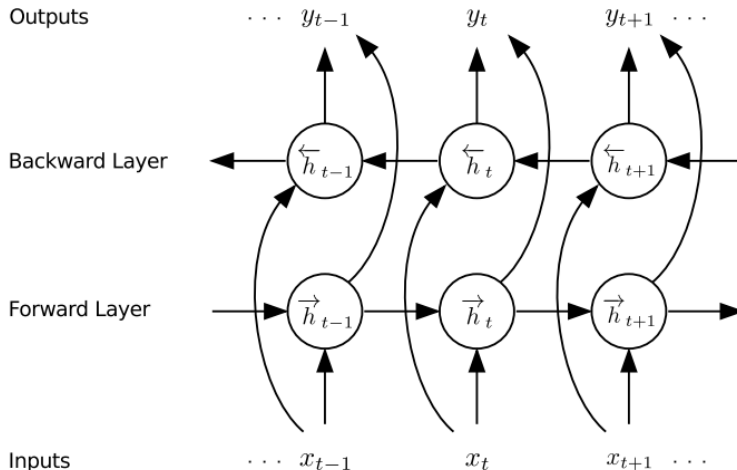


FIGURE 15: RNN bidirectionnel – Source : [Graves13b]

RNN bidirectionnel profond

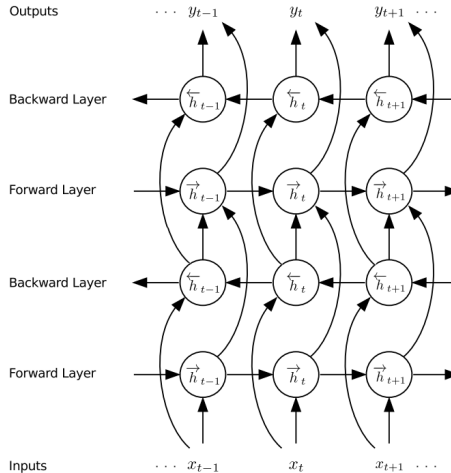


FIGURE 16: RNN bidirectionnel profond – Source : [Graves13b]

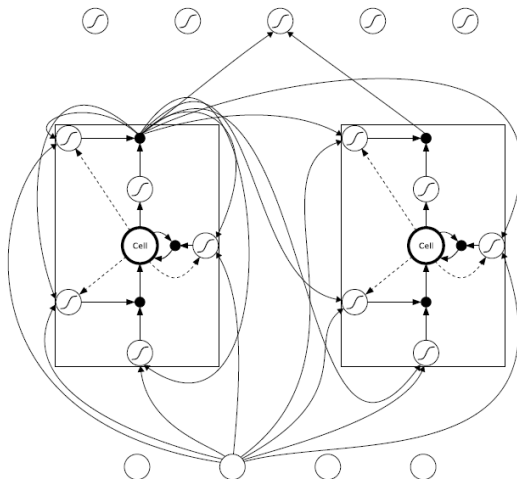


FIGURE 17: Architecture LSTM – Source : [Graves12]

LSTM bidirectionnel profond

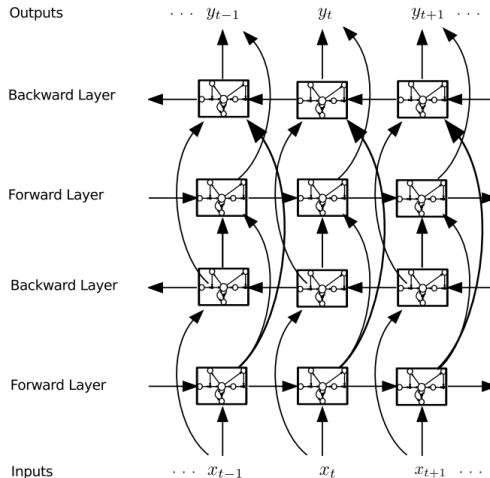


FIGURE 18: LSTM bidirectionnel profond (DB-LSTM) – Source : [Graves13b]

- Utilisation de méthodes d'apprentissage classique adaptées :
 - Real Time Recurrent Learning (RTRL) [**Robinson87**]
 - BackPropagation Through Time (BPTT) [**Williams95**]
- Possibilité d'approximer le gradient à chaque instant indépendamment des autres avec le RTRL
- Possibilité de calculer le gradient exact avec la BPTT [**Graves05b**]

- Prédiction de la structure de protéines [**Sonderby14**]
- Génération de musique [**Eck02**]
- Reconnaissance de l'écriture [**Graves09a ; Graves09b**]
 - Meilleure performance à l'état de l'art. Gain de 3 compétitions à l'ICDAR 2009.
- Reconnaissance de la parole [**Graves13a ; Graves13b**]
- Reconnaissance d'objets [**Ciresan11a**]
 - Gain de la compétition IJCNN 2011 en reconnaissance de panneaux de signalisation [**Ciresan11b**]
- Etc. (voir l'interview de Schmidhuber [**Angelica12**])






CONCLUSION

Conclusion






- Les réseaux de neurones sont apparus dans les années 1950-60, et les RNN dans les années 1980.
- Premiers résultats peu concluants
- Regain d'intérêt avec :
 - Augmentation de la puissance de calcul
 - Nouvelles méthodes (Deep Learning, LSTM, ...)
- Les LSTM offrent actuellement les meilleures performances à l'état de l'art sur de nombreux problèmes
- Méthodes jeunes, peu sorties des labos, recherche active

BIBLIOGRAPHIE







Bibliographie I

- 
 Amara D. ANGELICA et Jürgen SCHMIDHUBER. *How bio-inspired deep learning keeps winning competitions*. 2012. URL : <http://www.kurzweilai.net/how-bio-inspired-deep-learning-keeps-winning-competitions>.
- 
 Dan C CIREŞAN et al. “High-performance neural networks for visual object classification”. In : *arXiv preprint arXiv :1102.0183* (2011).
- 
 Dan CIREŞAN et al. “A committee of neural networks for traffic sign classification”. In : *Neural Networks (IJCNN), The 2011 International Joint Conference on*. IEEE. 2011, p. 1918–1921.
- 
 D. ECK et J. SCHMIDHUBER. “Finding temporal structure in music : blues improvisation with LSTM recurrent networks”. In : *Proceedings of the 2002 12th IEEE Workshop on Neural Networks for Signal Processing, 2002*. 2002, p. 747–756. DOI : 10.1109/NNSP.2002.1030094.
- 
 Felix A GERS, Jürgen SCHMIDHUBER et Fred CUMMINS. “Learning to forget : Continual prediction with LSTM”. In : *Neural computation* 12.10 (2000), p. 2451–2471. URL : <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.55.5709&rep=rep1&type=pdf>.







Bibliographie II

-  Felix A GERS et Jürgen SCHMIDHUBER. “LSTM recurrent networks learn simple context-free and context-sensitive languages”. In : *Neural Networks, IEEE Transactions on* 12.6 (2001), p. 1333–1340.
-  Felix A GERS, Nicol N SCHRAUDOLPH et Jürgen SCHMIDHUBER. “Learning precise timing with LSTM recurrent networks”. In : *The Journal of Machine Learning Research* 3 (2003), p. 115–143. URL : http://machinelearning.wustl.edu/mlpapers/paper_files/GersSS02.pdf.
-  Alex GRAVES, Santiago FERNÁNDEZ et Jürgen SCHMIDHUBER. “Bidirectional LSTM networks for improved phoneme classification and recognition”. In : *Artificial Neural Networks : Formal Models and Their Applications–ICANN 2005*. Springer, 2005, p. 799–804. URL : <http://www6.in.tum.de/pub/Main/Publications/Graves2005c.pdf>.
-  Alex GRAVES et Jürgen SCHMIDHUBER. “Framewise phoneme classification with bidirectional LSTM and other neural network architectures”. In : *Neural Networks* 18.5 (2005), p. 602–610. URL : http://www.cs.toronto.edu/~graves/nn_2005.pdf.
-  Alex GRAVES et al. “Connectionist temporal classification : labelling unsegmented sequence data with recurrent neural networks”. In : *Proceedings of the 23rd international conference on Machine learning*. ACM. 2006, p. 369–376.







Bibliographie III

- 
 Alex GRAVES et Jürgen SCHMIDHUBER. “Offline handwriting recognition with multidimensional recurrent neural networks”. In : *Advances in Neural Information Processing Systems*. 2009, p. 545–552.
- 
 Alex GRAVES et al. “A novel connectionist system for unconstrained handwriting recognition”. In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31.5 (2009), p. 855–868.
- 
 Alex GRAVES. *Supervised sequence labelling with recurrent neural networks*. T. 385. Springer, 2012.
- 
 Alex GRAVES, A-R MOHAMED et Geoffrey HINTON. “Speech recognition with deep recurrent neural networks”. In : *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE. 2013, p. 6645–6649.
- 
 Alex GRAVES, Navdeep JAITLEY et A-R MOHAMED. “Hybrid speech recognition with deep bidirectional LSTM”. In : *Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on*. IEEE. 2013, p. 273–278.
- 
 Donald HEBB. “0.(1949) The organization of behavior : A neuropsychological theory”. In : *New York* ().

Bibliographie IV

- 
 Sepp HOCHREITER. “Untersuchungen zu dynamischen neuronalen Netzen”. In : *Master’s thesis, Institut fur Informatik, Technische Universitat, Munchen* (1991).
- 
 Sepp HOCHREITER et Jürgen SCHMIDHUBER. “Long short-term memory”. In : *Neural computation* 9.8 (1997), p. 1735–1780. URL : http://deeplearning.cs.cmu.edu/pdfs/Hochreiter97_lstm.pdf.
- 
 Herbert JAEGER. “The “echo state” approach to analysing and training recurrent neural networks-with an erratum note”. In : *Bonn, Germany : German National Research Center for Information Technology GMD Technical Report* 148 (2001), p. 34.
- 
 Michael I JORDAN. “Serial Order : A Parallel Distributed Processing Approach.” In : (1986).
- 
 Wolfgang MAASS, Thomas NATSCHLÄGER et Henry MARKRAM. “Real-time computing without stable states : A new framework for neural computation based on perturbations”. In : *Neural computation* 14.11 (2002), p. 2531–2560.
- 
 Warren S McCULLOCH et Walter PITTS. “A logical calculus of the ideas immanent in nervous activity”. In : *The bulletin of mathematical biophysics* 5.4 (1943), p. 115–133.

Bibliographie V

- 
 Barak A PEARLMUTTER. “Learning state space trajectories in recurrent neural networks”.
 In : *Neural Computation* 1.2 (1989), p. 263–269.
- 
 AJ ROBINSON et Frank FALLSIDE. *The utility driven dynamic error propagation network*.
 University of Cambridge Department of Engineering, 1987.
- 
 F. ROSENBLATT. “The perceptron : A probabilistic model for information storage and
 organization in the brain”. In : *Psychological Review* 65.6 (1958), p. 386–408.
- 
 Søren Kaae SØNDERBY et Ole WINTHER. “Protein Secondary Structure Prediction with
 Long Short Term Memory Networks”. In : *arXiv :1412.7828 [cs, q-bio]* (déc. 2014). arXiv :
 1412.7828. URL : <http://arxiv.org/abs/1412.7828>.
- 
 Wern-Jun WANG, Yuan-Fu LIAO et Sin-Horng CHEN. “RNN-based prosodic modeling for
 mandarin speech and its application to speech-to-text conversion”. In : *Speech*
Communication 36.3 (2002), p. 247–265.
- 
 Ronald J WILLIAMS et David ZIPSER. “A learning algorithm for continually running fully
 recurrent neural networks”. In : *Neural computation* 1.2 (1989), p. 270–280.

Bibliographie VI



Ronald J WILLIAMS et David ZIPSER. “Gradient-based learning algorithms for recurrent networks and their computational complexity”. In : *Back-propagation : Theory, architectures and applications* (1995), p. 433–486.



Yunong ZHANG, Danchi JIANG et Jun WANG. “A recurrent neural network for solving Sylvester equation with time-varying coefficients”. In : *Neural Networks, IEEE Transactions on* 13.5 (2002), p. 1053–1063.