

Contexte scientifique : Depuis quelques années, la communauté scientifique des chercheurs du TAL s'efforce d'explorer la boîte noire que constituent les réseaux de neurones. Parallèlement aux travaux s'efforçant d'identifier les neurones et les couches de neurones impliqués pour des tâches comme la reconnaissance d'images (Montavon et al., 2018), plusieurs expériences ont été conduites pour essayer de comprendre comment les réseaux de neurones « apprennent », en particulier dans le domaine de la catégorisation linguistique (*pos-tagging*). Un atelier « blackbox » réunit depuis trois ans des chercheurs et des équipes lors de conférences telles que ACL, EMNLP ou COLING autour de ces problématiques. La majorité des expériences ne met pas en jeu le français et l'anglais et peu portent sur la traduction neuronale.

Tâche à réaliser: Réplication de quatre expériences avec des données français/anglais à partir du déploiement des systèmes décrits dans les articles. Analyse de ce que le RNN semble avoir « appris » avec des données annotées français/anglais. Par exemple, travaillant avec un réseau de neurones bi-LSTM préfigurateur d'OpenNMT, (Belinkov et al. 2017) semblent pouvoir établir une division du travail entre les couches inférieures (structures des mots) et couches supérieures (sens des mots) : l'analyse des résultats avec des données français/anglais confirme-t-elle ces conclusions ?

Données : Les données EUROPARL français/ anglais/allemand/espagnol dans leur version corrigée en https://pub.cl.uzh.ch/wiki/public/costep/start#known_errors. Le fichier contient plusieurs millions de tokens d'Europarl, commencer à l'année 2000 (fichiers de type 2000-07-05.xml)

références :

Belinkov, Y., Durrani, N., Dalvi, F., Sajjad, H., & Glass, J. (2017). What do neural machine translation models learn about morphology?. arXiv preprint arXiv:1704.03471.

<https://arxiv.org/pdf/1704.03471.pdf>

<https://github.com/boknilev/nmt-repr-analysis>

Poliak, A., Belinkov, Y., Glass, J., & Van Durme, B. (2018). On the evaluation of semantic phenomena in neural machine translation using natural language inference. arXiv preprint arXiv:1804.09779.

<https://www.aclweb.org/anthology/N18-2082.pdf>

<https://github.com/boknilev/nmt-repr-analysis>

Tal Linzen, Emmanuel Dupoux & Yoav Goldberg (2016). Assessing the ability of LSTMs to learn syntax-sensitive dependencies. Transactions of the Association for Computational Linguistics 4, 521–535. http://tallinzen.net/media/papers/linzen_dupoux_goldberg_2016_tacl.pdf

https://github.com/TalLinzen/rnn_agreement

K. Gulordava, P. Bojanowski, E. Grave, T. Linzen, M. Baroni. 2018. Colorless green recurrent networks dream hierarchically. Proceedings of NAACL. <https://arxiv.org/pdf/1803.11138>

<https://github.com/facebookresearch/colorlessgreenRNNs/tree/master/data>

LeDuolingo challenge <https://sharedtask.duolingo.com/>

BIBLIOGRAPHIE COMPLEMENTAIRE

Ballier, N., Amari, N., Merat, L. and Yunès, J.-B. (submitted) The Learnability of the Annotated Input in NMT (Replicating Vanmassenhove et al. 2018 with OpenNMT). LREC REPROLANG workshop, 8 p.

Montavon, G., Samek, W., and Müller, K.-R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73:1–15.