

Deep Learning avec R et H2O

Géraud Dugé de Bernonville

Valtech_

1, rue Dalayrac 31000 Toulouse

`geraud.dugedebernouville@gmail.com`

Mots clefs : Deep Learning, H2O, Big Data

Le Deep Learning fait de plus en plus le buzz ces derniers mois avec des utilisations dans la reconnaissance d'image[1], le traitement du langage[2] et le jeu de Go[3]. Ce regain d'intérêt pour les réseaux de neurones s'explique notamment par l'amélioration des algorithmes et également les progrès liés au matériel (notamment les GPU devenus de plus en plus puissants)[4].

Le langage R n'est pas en reste car il existe des outils permettant d'entraîner des modèles Deep Learning à partir de R. H2O[5] fait partie de ces outils. Développé en Java[6], H2O fournit des API permettant de s'intégrer à d'autres outils et langages tels que Python, Java et bien évidemment R[7]. H2O est un outil Open Source supporté par la société h2o.ai. Fondée par SriSatish Ambati en 2011, elle a depuis levé près de 30M\$ de fonds[8].

H2O propose de nombreuses fonctionnalités d'analyse de données et de machine learning dont:

- manipulation de données (filtre, conversions...)
- statistiques descriptives
- feature engineering (réduction de dimensions)
- machine learning (régressions, gradient boosting, random forest...)
- validation des modèles

H2O est une technologie basée sur des implémentations MapReduce capables de monter en charge et de s'exécuter sur des clusters (H2O, Spark, Hadoop). Cela permet ainsi d'apporter au langage R une scalabilité qui constitue souvent une pierre d'achoppement à son utilisation, et offre ainsi la possibilité d'utiliser R pour effectuer de l'analyse Big Data.

Au travers de ce *Lightning Talk*, nous allons présenter un aperçu des possibilités de H2O qui ont été mises en oeuvre pour la résolution d'un challenge de Data Science. Après une présentation des notions liées au Deep Learning, nous allons décrire les possibilités de H2O puis montrer un exemple d'utilisation de réseau de neurones sur un cas concret à l'aide du langage R.

Références

[1] Wu, R., Yan, S., Shan, Y., Dang, Q., & Sun, G. (2015). Deep image: Scaling up image recognition. arXiv preprint arXiv:1501.02876, 22, 388.

[2] Christopher Olah (2014). Deep Learning, NLP, and Representations. <http://colah.github.io/posts/2014-07-NLP-RNNs-Representations/>.

[3] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., ... & Dieleman, S. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484-489.

[4] Dettmers, T. (2015). Deep Learning in a Nutshell: History and Training. <https://devblogs.nvidia.com/parallelforall/deep-learning-nutshell-history-training/>

[5] Aiello, S., Eckstrand, E., Fu, A., Landry, M., Lanford, J., and Aboyoun, P. (Aug. 2015). Fast Scalable R with H2O. <http://h2o.ai/resources>.

[6] Candel, A., LeDell, E. (2016). Fast Scalable Machine Learning For Smarter Applications. <https://github.com/h2oai>.

[7] H2O.ai. (2016). Hadoop + R — H2O.ai (0xData) - Fast Scalable Machine Learning. <http://www.h2o.ai/product/integration/>.

[8] VB Profiles. (2016). Company H2O News, Employees and Funding Information, Mountain View, CA. <https://www.vbprofiles.com/companies/5460ec6db6a9dfdb5400c3e7#fundings>.