

Stage M2 : Évaluation de l'impact environnemental des méthodes de traitement automatique de la langue.

Mots-clés : traitement automatique de la langue, complexité algorithmique

Durée : 5 mois

Niveau : Master 2 (professionnel ou recherche), fin d'école d'ingénieur

Rémunération : Indemnité de stage, soit ~ 600 €/mois, indemnité de transport incluse

Lieu : Laboratoire LISN/LIMSI, campus de l'université Paris Saclay à Orsay

Contexte

De nombreux travaux en Traitement Automatique de la Langue (TAL) s'appuient sur des méthodes d'apprentissage. Ainsi, l'apprentissage profond offre des performances souvent intéressantes pour de nombreuses tâches d'analyse de textes. L'essor récent des méthodes neuronales donne lieu à une utilisation croissante de ressources numériques pour un large éventail de problèmes. Dans ce contexte, les méthodes symboliques ou méthodes d'apprentissage "classiques" sont délaissées alors qu'une comparaison systématique serait intéressante du point de vue scientifique, opérationnel et environnemental. En particulier, les méthodes neuronales ont un impact environnemental élevé qui ne cesse d'augmenter avec les années (Schwartz et al., 2019). Un exemple de travail pertinent en traitement automatique des langues est présenté par (Strubell et al. 2019), qui a étudié la consommation énergétique de l'apprentissage de plusieurs modèles. Schwartz et al. (2019) prône donc l'émergence de travaux en intelligence artificielle « verte » ou Green AI, en parallèle de travaux standards, dans lesquels l'efficacité des méthodes serait mise en valeur, c'est-à-dire la capacité à obtenir une performance avec un minimum de ressources. Ainsi il sera intéressant de proposer une comparaison détaillée de l'utilisation d'un large panel de méthodes de traitement automatique de la langue du point de vue de leur performance, de leur complexité algorithmique, du temps humain et machine requis pour les mettre en œuvre. Pour ce faire, il est nécessaire de s'intéresser à l'estimation de ces critères d'impact des méthodes numériques.

Objectifs du stage

Le stage a pour objectif de recenser et de caractériser les outils disponibles pour des travaux en intelligence artificielle verte en ce qui concerne l'estimation de l'impact environnemental des méthodes numériques. Les outils recensés seront mis en œuvre dans le cadre de l'application d'une méthode de traitement automatique de la langue (par exemple: classification de textes) afin de caractériser l'utilisation des outils sur le plan de la facilité de prise en main, qualité et niveau de détail des informations fournies sur les méthodes TAL, et tout autre critère d'évaluation pertinent.

Approche proposée

Le stage aura pour objet de réaliser une revue systématique des outils de mesure de l'impact environnemental des expériences informatiques. Une veille de la littérature récente montre que divers outils existent afin d'estimer l'impact des expériences informatiques. On recense notamment des outils en ligne (par exemple, Green Algorithms¹ et ML CO2 impact²) ou des outils à intégrer dans la mise en oeuvre des expériences (par exemple, "experiment impact tracker" (Henderson et al., 2020) et "carbon tracker" (Anthony et al., 2020)). Nous souhaitons recenser systématiquement les outils existants et les étudier afin de déterminer les mesures d'impact calculées, la facilité de mise en oeuvre, la portée d'utilisation possible.

Références

Anthony, L. F. W., Kanding, B., and Selvan, R. (2020). Carbontracker : Tracking and predicting the carbon footprint of training deep learning models. In ICML Workshop on "Challenges in Deploying and monitoring Machine Learning Systems".

Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., and Pineau, J. (2020). Towards the systematic reporting of the energy and carbon footprints of machine learning.

Schwartz, R., Dodge, J., Smith, N. A., and Etzioni, O. (2019). Green AI.

Strubell, E., Ganesh, A., and McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 3645–3650.

Compétences souhaitées:

Le stagiaire devra avoir de bonnes compétences en informatique. Des connaissances en traitement automatique de la langue seront particulièrement appréciées. Le contenu et l'ambition du stage pourront être modulés en fonction du niveau d'étude et de la durée du stage du/de la candidat.e.

Pour candidater :

Envoyer un CV, un relevé de notes récent ainsi que les coordonnées (nom, mail) d'au moins deux référent.e.s (professeur.e.s ou encadrant.e.s de précédents stages ou emplois pouvant attester de vos compétences) à Anne-Laure.Ligozat[at]limsi.fr et Aurelie.Neveol[at]limsi.fr

1 <http://www.green-algorithms.org/>

2 <https://mlco2.github.io/impact/>