
Convolution, Self-Attention et discours politique. Que faire des couches intermédiaires des réseaux de neurones profonds ?

Laurent Vanni^{*1} and Damon Mayaffre^{*†2}

¹Bases, Corpus, Langage (UMR 7320 - UCA / CNRS) – Bases, Corpus, Langage (UMR 7320 - UCA / CNRS) – France

²Bases, Corpus, Langage (UMR 7320 - UCA / CNRS) – Centre National de la Recherche Scientifique, Université Côte d’Azur – Laboratoire BCL - UMR 7320MSHS Sud-Est Pôle Universitaire Saint Jean d’Angely Bâtiment SJA 3 24 avenue des Diabls bleus, 06357 Nice Cedex 4, France

Résumé

Les grands modèles de langage (LLM) ont repoussé ces dernières années les limites du Traitement Automatique des Langues (TAL). Les architectures de type Transformer entraînées sur la quasi-totalité des productions écrites des humanités semblent aujourd’hui prêtes à répondre à toutes les questions envisageables des internautes les plus chevronnés. Cependant, si les sorties machines donnent l’illusion d’une compréhension des requêtes des utilisateurs, l’interprétation des résultats ne peut et ne doit être considérée par les SHS qu’au regard des programmes informatiques utilisés. Les travaux que nous proposons ont pour but d’éclairer la boîte noire de l’IA en utilisant l’ensemble de la chaîne de traitement des réseaux de neurones profonds, du corpus d’apprentissage à la prédiction (classification ou génération de textes). Chaque couche intermédiaire (ou couche cachée) est mobilisée pour extraire l’information apprise par les modèles qui conduit à une prédiction. Les représentations statistiques classiques des mots et/ou des textes se dotent ainsi de plus-values heuristiques nouvelles qui mêlent la vectorisation à grande échelle de l’information (*Embedding*), la détection locale d’association (syntagme) par *convolution* (Vanni et al. 2018) et la détection d’associations plus complexes et à distance (paradigme) par *Self-Attention* (Vanni et al. 2024). Appliqués à l’analyse de discours politique, ces travaux interdisciplinaires interrogent d’une part les linguistes ou les politologues sur le sens des textes et d’autre part les informaticiens sur le sens des hyperparamètres du *deep learning*. Avec une approche dirigée par le corpus – *corpus-driven* – le retour au texte permet une interprétation des marqueurs repérés par le modèle et conduit à une herméneutique des sorties logicielles (Rastier 2011). La prédiction de l’I.A. n’est alors plus une finalité mais un point d’étape nécessaire pour une exploration nouvelle des corpus numériques par les SHS.

Mots-Clés: Deep Learning, Linguistique, Texte, Convolution, Self, Attention, Transformer, Explicabilité

*Intervenant

†Auteur correspondant: damon.mayaffre@univ-cotedazur.fr