

Classification non supervisée de documents à partir des modèles Transformeurs

Mira Ait-Saada^{*,**}, François Role^{*}, Mohamed Nadif^{*}

^{*}Université de Paris, CNRS, Centre Borelli UMR9010, 75006 Paris

^{**}Caisse des Dépôts et Consignations, Datalab, 75013, Paris

<prénom.nom>@u-paris.fr

Résumé. Les modèles *Transformeurs* créent des plongements différents pour la même entrée, un à chaque couche de leur architecture. Diverses études ont déjà tenté d'identifier les plongements qui contribuent le plus au succès des tâches de classification supervisée. En revanche, la même analyse des performances n'a pas encore été réalisée dans le cadre non supervisé. Dans cet article, nous évaluons l'efficacité des modèles *Transformeurs* sur l'importante tâche de classification non supervisée de documents. Nous présentons une approche *clustering ensemble* qui exploite toutes les couches du réseau. Des expériences menées sur des ensembles de données réels avec différents modèles montrent l'efficacité de la méthode proposée par rapport à plusieurs stratégies habituellement utilisées. Cet article est une restitution du papier (Ait-Saada et al., 2021).

1 Introduction

Un modèle Transformeur produit plusieurs représentations pour chaque mot (une à chaque couche de l'architecture du réseau) et des études dans le domaine de l'apprentissage supervisé ont tenté de déterminer le type d'information capturé par les différentes couches. Par exemple, dans (Tenney et al., 2019), en utilisant des plongements issus de modèles Transformeur pré-entraînés en entrée d'une suite de tâches de traitement automatique de la langue, les auteurs ont observé que les premières couches encodent la plupart des phénomènes syntaxiques locaux tandis que des aspects sémantiques plus complexes apparaissent aux couches supérieures. D'autre part, s'intéressant plus spécifiquement aux capacités de généralisation des plongements de mots contextualisés (comprenant les algorithmes ELMo, BERT et OpenAI *Transformer*), Liu et al. (2019a) ont observé que les couches intermédiaires des Transformeurs présentent une meilleure transférabilité, tandis que Hao et al. (2019) ont observé que les premières couches de BERT-large sont plus stables et varient moins d'une tâche à l'autre lorsque le modèle est ré-entraîné sur différentes tâches et en ont conclu qu'elles étaient plus transférables que les couches supérieures. Dans un autre axe de recherche, certaines études se sont concentrées sur l'impact du *fine-tuning* ou ré-entraînement des modèles Transformeurs, et ont vérifié expérimentalement que plus on se rapproche de la dernière couche, plus les représentations deviennent spécifiques à la tâche apprise par le modèle (van Aken et al., 2019). Le principal point à retenir de ces études est que les représentations fournies par les différentes couches capturent clairement des informations différentes, conduisant ainsi à des résultats très différents lorsqu'elles sont