

Comment recentrer un géant sur une tâche spécifique : les grands modèles de langage et la Reconnaissance d'Entités Nommées

Anonymous ACL submission

1 Introduction

L'adaptabilité supérieure des grands modèles de langage est à double tranchant : lorsqu'ils effectuent des tâches spécifiques plutôt que générales, notamment lorsque lesdites tâches relèvent d'un domaine défini nécessitant de mobiliser des compétences précises ou d'effectuer une réflexion approfondie antérieurement à leur complétion, les résultats produits par ces modèles tendent à perdre en qualité (Wang et al., 2024). Cette problématique a déjà trouvé une forme de résolution du côté des acteurs techniques de la création de modèles : en spécialisant les modèles lors d'un processus de fine-tuning, il est en effet possible de produire des modèles qui soient plus adaptés à une tâche spécifique (Chung et al., 2022).

Toutefois, cette solution, en plus de demander le déploiement d'effort et de ressources supplémentaires pour répondre à des besoins très ciblés, n'est pas à la portée des acteurs de la recherche en science humaine et sociale qui n'auraient pas reçu de formation adéquate. L'enjeu principal consiste dès lors à rendre d'une part les humanités numériques plus accessibles, et d'autre part à encourager les méthodes de recherche traditionnelles à se compléter d'une analyse relevant des humanités numériques, si pertinent. Cette présentation se propose donc de présenter des solutions demandant de faibles compétences techniques pour mobiliser les grands modèles de langage dans un projet de recherche en sciences humaines et sociales.

En abordant la question de la Reconnaissance d'Entité Nommées [REN], une tâche spécifique et très utilisée dans les projets de sciences humaines et sociales, la présente entend illustrer les démarches proposées et justifier de leur pertinence. En effet, la REN est une tâche dont les paramètres et la nature exacte varieront selon les besoins du projet et la position des annotateurs, nécessitant la création de guides d'annotation avant chaque cam-

pagne d'annotation (Ehrmann, 2008). L'utilisation d'un grand modèle de langage ne peut alors répondre qu'aux besoins les plus basiques et les plus généraux, et deviendra de moins en moins pertinente avec la complexification des instructions d'annotation associées au projet. Il s'agit donc là d'un exemple particulièrement adapté à l'étude de la problématique annoncée.

2 Objectifs et questions de recherche

La présentation consistera en une évolution progressive du cas le plus au cas le plus complexe observé lors de l'utilisation de grands modèles de langage en REN : nous identifierons alors la problématique spécifique et des manières de la résoudre ou de compenser, dans la mesure du possible. Ainsi, nous traiterons dans cet ordre du cas où les modèles sont capables de produire des résultats excellents sur une tâche spécifique : en le considérant comme un outil de travail, il s'agira alors d'expliquer la meilleure façon de sélectionner le ou les modèles qui seront appliqués à la tâche et d'explicitier la façon dont un prompt peut être modifié de façon à améliorer les résultats produits ; du cas où le modèle produit des résultats mitigés, mais exploitables : en le considérant comme un annotateur de faible fiabilité, nous expliquerons comment l'intégrer néanmoins à une équipe d'annotation, et comment tirer profit du travail humain sur cette même tâche pour améliorer les résultats du modèle. Enfin, dans le cas où le modèle est complètement incompetent sur la tâche donnée, nous présenterons des utilisations alternatives : en le prenant tel qu'il est, c'est-à-dire un générateur à l'état de l'art, comment le mobiliser pour faire de l'augmentation de données génératives et renforcer la pertinence de son corpus pour l'annotation humaine.

3 Méthodologie proposée

La présentation consistera en une évolution progressive du cas le plus au cas le plus complexe observé lors de l'utilisation de grands modèles de langage en REN : nous identifierons alors la problématique spécifique et des manières de la résoudre ou de compenser, dans la mesure du possible.

Ainsi, nous traiterons dans cet ordre :

- du cas où les modèles sont capables de produire des résultats excellents sur une tâche spécifique : en le considérant comme un outil de travail, il s'agira alors d'expliquer la meilleure façon de sélectionner le ou les modèles qui seront appliqués à la tâche et d'explicitier la façon dont un prompt peut être modifié de façon à améliorer les résultats produits ;
- du cas où le modèle produit des résultats mitigés, mais exploitables : en le considérant comme un annotateur de faible fiabilité, nous expliquerons comment l'intégrer néanmoins à une équipe d'annotation, et comment tirer profit du travail humain sur cette même tâche pour améliorer les résultats du modèle ;
- enfin, du cas où le modèle est complètement incompetent sur la tâche donnée, nous présenterons des utilisations alternatives : en le prenant tel qu'il est, c'est-à-dire un générateur à l'état de l'art, comment le mobiliser pour faire de l'augmentation de données génératives et renforcer la pertinence de son corpus pour l'annotation humaine.

4 Revue de la littérature

Il existe déjà des modèles adaptés à la reconnaissance d'entité nommées : dans le cas du modèle NuNER (Bogdanov et al., 2024), le modèle peut créer de nouveaux concepts selon les besoins du texte. Cette capacité est toutefois un point négatif dans le cas, très large, où un projet ne demande l'annotation que de catégories spécifiques : en outre, la répartition des catégories n'est pas consistante et bien des concepts pourraient être regroupés dans une seule catégorie afin de gagner en pertinence et en visibilité. Les ambiguïtés propres aux entités nommées (Ehrmann, 2008) sont également sources d'erreur fréquentes pour les modèles.

L'article (Ashok and Lipton, 2023) donne des recommandations pour le prompting en matière de reconnaissances d'entités nommées.

En fournissant au modèle des *negative samples* représentatifs (Robinson et al., 2021), il est possible d'améliorer ses performances : toute campagne d'annotation comportant son lot d'erreur et de corrections à apporter, on peut imaginer utiliser ces erreurs humaines passées pour améliorer le modèle.

Il est possible de faire appel à des ressources et des connaissances extérieures pour augmenter la fiabilité du modèle sur un corpus spécifique (Bian et al., 2023).

L'annotation en crowdsourcing étant un procédé dont la fiabilité a été démontrée (Nowak and Rüger, 2010), on peut aussi imaginer d'intégrer le modèle en tant qu'annotateur parmi des annotateurs humains afin d'augmenter les retours possibles.

Enfin, les grands modèles de langage peuvent être mobilisés dans l'augmentation de données, tout particulièrement pour répondre aux problèmes de *class imbalance* (Suhaeni and Yong, 2023).

5 Résultats anticipés

Nous espérons que la présente permettra l'exposition de méthodes innovantes dans le domaine des sciences humaines et sociales, tout en facilitant la prise en main des grands modèles de langage en tant qu'outil de travail.

References

- Dhananjay Ashok and Zachary C. Lipton. 2023. [Promptner: Prompting for named entity recognition](#). *Preprint*, arXiv:2305.15444.
- Junyi Bian, Jiaxuan Zheng, Yuyi Zhang, and Shanfeng Zhu. 2023. [Inspire the large language model by external knowledge on biomedical named entity recognition](#). *Preprint*, arXiv:2309.12278.
- Sergei Bogdanov, Alexandre Constantin, Timothée Bernard, Benoit Crabbé, and Etienne Bernard. 2024. [Nuner: Entity recognition encoder pre-training via llm-annotated data](#). *Preprint*, arXiv:2402.15343.
- Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Yunxuan Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, Albert Webson, Shixiang Shane Gu, Zhuyun Dai, Mirac Suzgun, Xinyun Chen, Aakanksha Chowdhery, Alex Castro-Ros, Marie Pellat, Kevin Robinson, Dasha Valter, Sharan Narang, Gaurav Mishra, Adams Yu, Vincent Zhao, Yanping Huang, Andrew Dai, Hongkun Yu, Slav Petrov, Ed H. Chi, Jeff Dean, Jacob Devlin, Adam Roberts, Denny Zhou, Quoc V. Le, and Jason Wei. 2022. [Scaling instruction-finetuned language models](#). *Preprint*, arXiv:2210.11416.
- Maud Ehrmann. 2008. *Les Entités Nommées, de la linguistique au TAL : Statut théorique et méthodes*

- 176 *de désambiguïsation*. Ph.D. thesis, Paris Diderot
177 University,.
- 178 Stefanie Nowak and Stefan Rüger. 2010. [How reliable](#)
179 [are annotations via crowdsourcing: a study about](#)
180 [inter-annotator agreement for multi-label image an-](#)
181 [notation](#). In *Proceedings of the International Confer-*
182 *ence on Multimedia Information Retrieval, MIR '10*,
183 page 557–566, New York, NY, USA. Association for
184 Computing Machinery.
- 185 Joshua Robinson, Ching-Yao Chuang, Suvrit Sra, and
186 Stefanie Jegelka. 2021. [Contrastive learning with](#)
187 [hard negative samples](#). *Preprint*, arXiv:2010.04592.
- 188 Cici Suhaeni and Hwan-Seung Yong. 2023. [Mitigating](#)
189 [class imbalance in sentiment analysis through gpt-](#)
190 [3-generated synthetic sentences](#). *Applied Sciences*,
191 13(17).
- 192 Zhihu Wang, Shiwan Zhao, Yu Wang, Heyuan Huang,
193 Jiaxin Shi, Sitao Xie, Zhixing Wang, Yubo Zhang,
194 Hongyan Li, and Junchi Yan. 2024. [Re-task: Revis-](#)
195 [iting llm tasks from capability, skill, and knowledge](#)
196 [perspectives](#). *Preprint*, arXiv:2408.06904.