

PromptORE – Vers l’Extraction de Relations non-supervisée

Pierre-Yves Genest^{1,2} Pierre-Edouard Portier² Előd Egyed-Zsigmond²
Laurent-Walter Goix³ *

(1) Alteca, 88 Boulevard des Belges, 69006 Lyon, France

(2) Univ Lyon, INSA Lyon, LIRIS, CNRS UMR5205, 20 Avenue Einstein, 69621 Villeurbanne, France

(3) Nokia, 37 Quai du Président Roosevelt, 92130 Issy-les-Moulineaux, France

pygenest@alteca.fr, pierre-edouard.portier@insa-lyon.fr,
elod.egyed-zsigmond@insa-lyon.fr, laurent-walter.goix@nokia.com

RÉSUMÉ

L’extraction de relations non-supervisée vise à identifier les relations qui lient les entités dans un texte sans utiliser de données annotées pendant l’entraînement. Cette tâche est utile en monde ouvert, où les types de relations et leur nombre sont inconnus. Bien que des modèles récents obtiennent des résultats prometteurs, ils dépendent fortement d’hyper-paramètres dont l’ajustement nécessite des données annotées, signifiant que ces modèles ne sont pas complètement non-supervisés. Pour résoudre ce problème, nous proposons PromptORE, à notre connaissance le premier modèle d’extraction de relations non-supervisé qui ne nécessite pas d’ajustement d’hyper-paramètres. Pour cela, nous adaptons le principe du prompt-tuning pour fonctionner sans supervision. Les résultats montrent que PromptORE surpasse largement les méthodes à l’état de l’art, avec un gain relatif de 20 – 40% en B³, V-measure et ARI. Le code source est accessible ¹.

ABSTRACT

PromptORE – A Novel Approach Towards Fully Unsupervised Relation Extraction

Unsupervised relation extraction aims to identify relations between entities in text, without having access to labeled data during training. This setting is particularly relevant for open-domain relation extraction where relation types and the number of relations are *a priori* unknown. Although recent approaches achieve promising results, they heavily depend on hyperparameters whose tuning requires labeled data, meaning they are not fully unsupervised. To mitigate this issue, we propose PromptORE, to the best of our knowledge, the first unsupervised relation extraction model that does not need hyperparameter tuning. We adapt the novel prompt-tuning paradigm to work in an unsupervised setting. Results shows that PromptORE consistently outperforms state-of-the-art models with a relative gain of 20 – 40% in B³, V-measure and ARI. The source code is available ¹.

MOTS-CLÉS : extraction de relation non-supervisée ; extraction de relation ouverte ; prompt-tuning.

KEYWORDS: unsupervised relation extraction ; open relation extraction ; prompt-tuning.

*. Recherche réalisée lorsque l’auteur travaillait chez Alteca.

1. <https://github.com/alteca/PromptORE>.

1 Introduction

L'Extraction de Relations Non-Supervisée (ER-NS) vise à identifier les relations qui lient les entités dans un texte, sans avoir accès à des données annotées pendant l'entraînement. Cette tâche est particulièrement utile sur des domaines très spécifiques où aucun jeu de données annoté n'est disponible, ou en monde ouvert où les types de relations et leur nombre ne sont pas connus à l'avance. Plusieurs approches récentes s'intéressent à cette thématique et obtiennent des résultats prometteurs (Simon *et al.*, 2019; Tran *et al.*, 2020; Hu *et al.*, 2020), mais elles dépendent toutes de nombreux hyper-paramètres, dont l'ajustement nécessite des données annotées.

Pour répondre à ce problème, nous proposons **PromptORE**, un modèle d'*Extraction de Relations en monde Ouvert basée sur des Prompts*², que nous avons présenté lors de la conférence CIKM'22 (Genest *et al.*, 2022). Nous utilisons et modifions la nouvelle technique du *prompt-tuning* pour fonctionner en mode non-supervisé. Nous l'utilisons pour calculer une représentation vectorielle de chaque relation (*relation embedding*). Nous effectuons ensuite un clustering sur ces embeddings pour identifier des groupes de relations similaires, et effectuer des prédictions. À notre connaissance, PromptORE est le premier modèle d'ER-NS qui ne nécessite pas d'ajustement d'hyper-paramètres. Les résultats obtenus sur des datasets de domaines général et spécifique montrent que PromptORE surpasse largement les méthodes à l'état de l'art, avec un gain relatif de 20 – 40% en B³, V-mesure et ARI, tout en étant plus simple. Une analyse qualitative montre aussi la capacité de PromptORE à extraire des clusters sémantiquement cohérents et très proches des vraies relations.

2 État de l'art

L'extraction de relations non-supervisée (ER-NS) vise à identifier la relation binaire r qui lie deux entités e_1 et e_2 dans le contexte d'une phrase S ³, sans utiliser un jeu de données annoté. Le plus souvent, il s'agit de regrouper dans un cluster toutes les instances (S, e_1, e_2) qui expriment la même relation r . Les approches récentes s'appuient sur des modèles de langage, notamment BERT (Devlin *et al.*, 2019), et suivent un processus en deux étapes (Hu *et al.*, 2020; Simon *et al.*, 2019; Marcheggiani & Titov, 2016) : 1. encodage des instances, c'est-à-dire représenter chaque relation par un vecteur appelé *plongement de relation*, et 2. prédiction de la relation en utilisant le plongement. Simon *et al.* (2019); Marcheggiani & Titov (2016) utilisent une approche générative basée sur les auto-encodeurs variationnels. SelfORE (Hu *et al.*, 2020) utilise l'apprentissage auto-supervisé, en générant des pseudo-labels avec un clustering et en les prédisant avec un classificateur pour affiner les paramètres de BERT (*fine-tuning*). Cependant, Tran *et al.* (2020) montrent avec EType+ qu'un modèle très simple, raisonnant sur les types d'entités, obtient de meilleures performances que ces approches complexes.

Bien que tous ces modèles extraient des relations sans utiliser de données annotées, nous ne pensons pas qu'ils soient réellement non-supervisés. En effet, ils possèdent beaucoup d'hyper-paramètres comme le nombre de clusters, le nombre d'époques, d'itérations, le taux d'apprentissage, l'early-stopping, ou la régularisation. Ajuster ces hyper-paramètres est critique pour les performances des modèles, et les auteurs ne décrivent aucune méthode pour les ajuster automatiquement et sans données

2. *Prompt* peut être traduit par réplique ou requête.

3. Les modèles d'EN-RS se concentrent sur l'extraction dans des phrases, bien que l'extraction dans des documents soit plus générale.

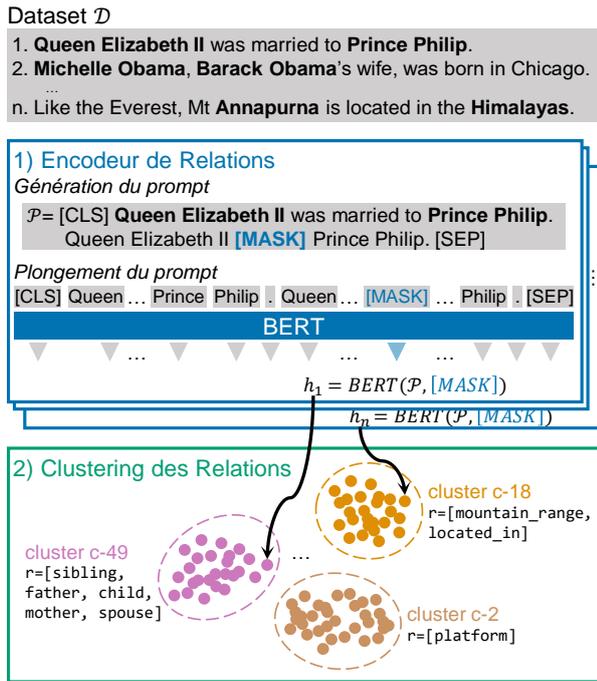


FIGURE 1 – Architecture de PromptORE.

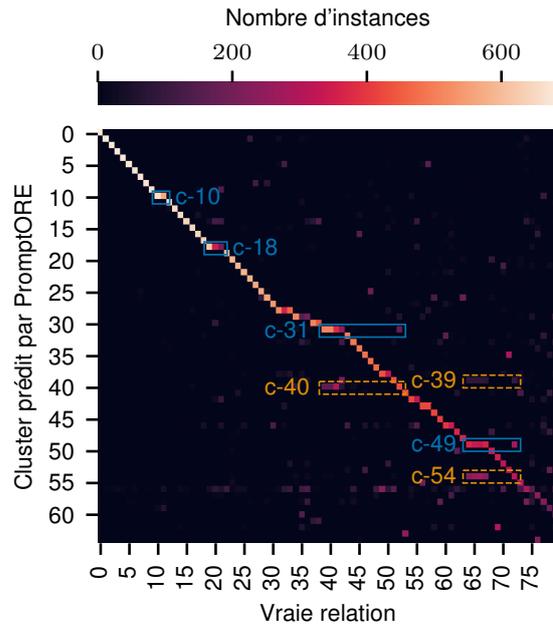


FIGURE 2 – Matrice de confusion entre les relations et les clusters de PromptORE sur FewRel.

annotées. De fait, les approches existantes ne sont pas utilisables dans un contexte non-supervisé réel.

3 PromptORE

C’est pour répondre à ce problème que nous proposons PromptORE. Il a pour objectif d’extraire la relation binaire r entre deux entités $e1$ et $e2$ déjà extraites dans une phrase S , sans utiliser de données annotées pendant l’apprentissage et l’ajustement des hyper-paramètres. Nous supposons avoir accès à un dataset \mathcal{D} contenant n instances (voir Figure 1). Chaque instance est décrite par un triplet $(S, e1, e2)$, avec S le texte de l’instance et $e1$ et $e2$ les deux entités. Nous supposons $e1$ et $e2$ extraits, et qu’il existe une relation r qui relie $e1$ et $e2$ dans le contexte de S (hypothèse habituelle de l’ER-NS). \mathcal{R} est l’ensemble des k types de relations de \mathcal{D} . Nous ne connaissons pas k , ni n’avons d’information sur les relations (label, description, types d’entités, etc.). Comme SelfORE (Hu *et al.*, 2020), PromptORE est composé de deux modules principaux (voir Figure 1) : l’encodeur de relations, et le clustering des relations.

Encodeur de relations Ce module calcule une représentation vectorielle de la relation (*relation embedding*) qui lie $e1$ à $e2$ dans le contexte de S . En d’autres termes, l’encodeur de relation abstrait le texte, pour fournir un plongement qui est facile à comparer grâce à une métrique (distance euclidienne). L’objectif est d’avoir un plongement représentatif de la relation sous-jacente : si deux instances ont des plongements proches, alors elles manifestent vraisemblablement la même relation r . Pour notre encodeur, nous utilisons le principe du *prompt-tuning* avec le modèle de langage BERT, que nous adaptons pour fonctionner dans un cadre non-supervisé⁴ :

4. Le prompt-tuning a déjà été utilisé pour l’extraction de relation, mais dans un cadre supervisé (e.g., Lv *et al.* (2022)).

| Dataset | | Modèle | B ³ (F1) | V (F1) | ARI |
|---|----------------|------------------------------------|---------------------|-------------|-------------|
| FewRel (Han <i>et al.</i> , 2018) | $k = 80$ | EType+ (Tran <i>et al.</i> , 2020) | 13.7 | 47.9 | 8.4 |
| | $\hat{k} = 65$ | SelfORE (Hu <i>et al.</i> , 2020) | 29.2 | 53.2 | 24.4 |
| | | PromptORE | 49.5 | 71.2 | 42.2 |
| FewRel PubMed (Gao <i>et al.</i> , 2019) | $k = 10$ | EType+ (Tran <i>et al.</i> , 2020) | 18.1 | 0 | 0 |
| | $\hat{k} = 10$ | SelfORE (Hu <i>et al.</i> , 2020) | 59.3 | 63.4 | 45.4 |
| | | PromptORE | 77.4 | 81.1 | 73.8 |

TABLE 1 – Résultats de PromptORE et de modèles à l’état de l’art sur deux jeux de données. EType+ et SelfORE ont accès au nombre de relations k , PromptORE l’estime (\hat{k}).

1. Pour une instance (\mathbf{S} , $\mathbf{e1}$, $\mathbf{e2}$), déterminer un prompt \mathcal{P} , qui est une séquence de tokens incluant un token [MASK]. Dans un cadre de *prompt-tuning* classique, l’objectif est d’optimiser la formulation de \mathcal{P} pour améliorer les performances. Sans accès à des données annotées, cela nous est impossible, aussi nous proposons d’utiliser la template la plus simple possible :

$$\mathcal{P}(\mathbf{S}, \mathbf{e1}, \mathbf{e2}) = "[CLS] \ \mathbf{S} \ \mathbf{e1} \ [MASK] \ \mathbf{e2}. [SEP]". \quad (1)$$

Un exemple de prompt est disponible sur la Figure 1.

2. Prédire le plongement h du token [MASK] dans le contexte de \mathcal{P} en utilisant BERT : $h = \text{BERT}(\mathcal{P}, [\text{MASK}])$. h est le *relation embedding*, le plongement de la relation.

Les approches précédentes d’ER-NS utilisent la technique de la représentation des paires d’entités pour l’encodeur (Hu *et al.*, 2020; Simon *et al.*, 2019; Marcheggiani & Titov, 2016) : elles calculent un plongement pour $\mathbf{e1}$ et $\mathbf{e2}$, grâce à l’ajout de tokens virtuels, et concatènent les deux plongements pour obtenir le *relation embedding*. Cette approche nécessite de fine-tuner BERT, ce qui est la source de beaucoup d’hyper-paramètres. Notre implémentation basée sur le prompt-tuning, ne nécessite pas de fine-tuning, supprimant de fait tous ces hyper-paramètres.

Clustering des relations Pour identifier les groupes d’instances manifestant la même relation r , nous calculons un clustering avec les k-moyennes (Lloyd, 1957) sur les plongements h de toutes les instances de \mathcal{D} (voir Figure 1). Pour estimer le nombre de relations k , nous proposons d’utiliser l’*elbow rule* (Thorndike, 1953) en conjonction avec le coefficient silhouette (Rousseeuw, 1987). Elle permet de trouver un nombre de clusters \hat{k} qui est un compromis entre la simplicité du clustering et la valeur du coefficient silhouette.

4 Expériences

Jeux de données, Métriques & Comparaisons Nous utilisons deux jeux de données : FewRel (Han *et al.*, 2018), de domaine général (articles Wikipedia) et contenant 56 000 instances avec 80 types de relations ; et FewRel PubMed (Gao *et al.*, 2019), centré sur le domaine biomédical, avec 10 types de relations différents. Étant donné que nous calculons un clustering comme les travaux précédents, nous évaluons PromptORE avec les métriques B³ (Bagga & Baldwin, 1998), V-mesure (V) (Rosenberg & Hirschberg, 2007) et Adjusted Rand Index (ARI) (Hubert & Arabie, 1985). Nous

nous comparons aux deux approches à l'état de l'art SelfORE (Hu *et al.*, 2020) et EType+ (Tran *et al.*, 2020). Ces deux approches sont entraînées avec les valeurs des hyper-paramètres déterminées par leurs auteurs respectifs. En particulier, ils ont besoin de connaître le nombre de relations k , ce qui n'est pas le cas de PromptORE.

Résultats quantitatifs⁵ Les résultats sont disponibles sur la Table 1. De manière générale, PromptORE est bien plus performant que SelfORE et EType+ sur les trois métriques : l'écart est de $\approx 19\%$ en B^3 , $\approx 18\%$ en V-measure et $\approx 23\%$ en ARI. PromptORE ne voit pas ses performances réduites sur FewRel PubMed (domaine biomédical), ce qui démontre sa flexibilité et son adaptabilité à d'autres domaines. SelfORE aussi n'est pas impacté, par contre EType+ ne prédit qu'une seule relation pour FewRel PubMed (ARI et V-measure à 0), montrant la limite du raisonnement sur les types d'entités pour des domaines très spécifiques. Les performances de PromptORE sont d'autant plus impressionnantes qu'il n'a pas accès à k , contrairement à SelfORE et EType+. PromptORE démontre qu'il est possible d'extraire des relations en mode non-supervisé sans hyper-paramètres à ajuster, tout en ayant de bien meilleures performances que les modèles à l'état de l'art.

Analyse qualitative⁵ Sur la Table 1, nous remarquons que l'estimation \hat{k} par l'elbow rule est correcte pour FewRel PubMed, mais qu'elle est sous-estimée pour FewRel : 65 clusters contre 80 relations. Cela signifie que certains clusters doivent être impurs, c'est-à-dire contenir les instances de plusieurs relations à la fois. La matrice de confusion de PromptORE évalué sur FewRel est disponible en Figure 2. Nous voyons que certains clusters sont impurs (e.g., c-10, c-18, c-31, c-49), mais dans l'ensemble nous observons une diagonale claire : PromptORE extrait précisément la grande majorité des relations, malgré son fonctionnement non-supervisé. La plupart des clusters sont relativement complets : peu de clusters partagent les mêmes relations (à l'exception de c-31, c-40 et c-39, c-49, c-54). En observant les clusters contenant plusieurs relations, nous remarquons qu'ils sont sémantiquement cohérents. Pour ne citer que deux exemples :

- c-10. Deux relations : `language_of_film_or_tv_show` et `language_of_work_or_name`. Ce sont des relations centrées sur le domaine du langage d'une oeuvre artistique.
- c-49. Cinq relations : `sibling`, `father`, `child`, `mother` et `spouse`. Relations autour des liens familiaux.

Du point de vue de FewRel, ces clusters sont impurs et incorrects, mais d'un point de vue qualitatif les relations extraites sont sémantiquement proches et cohérentes.

5 Conclusion

Nous avons introduit PromptORE, un modèle d'extraction de relations non-supervisé sans hyper-paramètres à ajuster, qui s'appuie sur la nouvelle méthode du prompt-tuning. Bien que très simple, PromptORE surpasse largement les approches à l'état de l'art. Pour le futur, nous envisageons de *fermer la boucle* de la connaissance, c'est-à-dire bénéficier de la connaissance extraite par PromptORE pour améliorer itérativement les performances de l'extraction.

5. Une évaluation plus détaillée de PromptORE est accessible sur (Genest *et al.*, 2022), contenant plus de datasets, d'approches comparées, et plusieurs formulations pour le prompt \mathcal{P} .

Remerciements

Ce travail est soutenu par Alteca et l'Association Nationale de la Recherche et de la Technologie (ANRT) par l'intermédiaire de la thèse CIFRE n°2021/0851. Nous remercions les relecteurs anonymes pour leurs remarques instructives et commentaires pertinents.

Références

- BAGGA A. & BALDWIN B. (1998). Algorithms for scoring coreference chain. In *Proceedings of the 1st International Conference on Language Resources and Evaluation Workshop on Linguistics Coreference*, p. 563–566, Granada, Spain : European Language Resources Association.
- DEVLIN J., CHANG M. W., LEE K. & TOUTANOVA K. (2019). BERT : Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, volume 1, p. 4171–4186, Stroudsburg, PA, USA : Association for Computational Linguistics. DOI : [10.18653/V1/N19-1423](https://doi.org/10.18653/V1/N19-1423).
- GAO T., HAN X., ZHU H., LIU Z., LI P., SUN M. & ZHOU J. (2019). Fewrel 2.0 : Towards more challenging few-shot relation classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing*, p. 6250–6255, Hong Kong, China : Association for Computational Linguistics. DOI : [10.18653/v1/d19-1649](https://doi.org/10.18653/v1/d19-1649).
- GENEST P.-Y., PORTIER P.-E., EGYED-ZSIGMOND E. & GOIX L.-W. (2022). PromptORE - A Novel Approach Towards Fully Unsupervised Relation Extraction. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, p. 11, Atlanta, USA : ACM. DOI : [10.1145/3511808.3557422](https://doi.org/10.1145/3511808.3557422).
- HAN X., ZHU H., YU P., WANG Z., YAO Y., LIU Z. & SUN M. (2018). Fewrel : A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, p. 4803–4809, Brussels, Belgium : Association for Computational Linguistics. DOI : [10.18653/v1/d18-1514](https://doi.org/10.18653/v1/d18-1514).
- HU X., WEN L., XU Y., ZHANG C. & YU P. S. (2020). SelfORE : Self-supervised relational feature learning for open relation extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, p. 3673–3682, Online : Association for Computational Linguistics. DOI : [10.18653/v1/2020.emnlp-main.299](https://doi.org/10.18653/v1/2020.emnlp-main.299).
- HUBERT L. & ARABIE P. (1985). Comparing partitions. *Journal of Classification*, **2**(1), 193–218. Publisher : Springer, DOI : [10.1007/BF01908075](https://doi.org/10.1007/BF01908075).
- LLOYD S. P. (1957). Least squares quantization in PCM. *Technical Report RR-5497*.
- LV B., JIN L., ZHANG Y., WANG H., LI X. & GUO Z. (2022). Commonsense Knowledge-Aware Prompt Tuning for Few-Shot NOTA Relation Classification. *Applied Sciences*, **12**(4), 2185–2185. Publisher : Multidisciplinary Digital Publishing Institute, DOI : [10.3390/app12042185](https://doi.org/10.3390/app12042185).
- MARCHEGGIANI D. & TITOV I. (2016). Discrete-State Variational Autoencoders for Joint Discovery and Factorization of Relations. *Transactions of the Association for Computational Linguistics*, **4**, 231–244. Publisher : MIT Press - Journals, DOI : [10.1162/tacl_a_00095](https://doi.org/10.1162/tacl_a_00095).

- ROSENBERG A. & HIRSCHBERG J. (2007). V-Measure : A conditional entropy-based external cluster evaluation measure. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, p. 410–420, Prague, Czech Republic : Association for Computational Linguistics.
- ROUSSEEUW P. J. (1987). Silhouettes : A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, **20**(C), 53–65. Publisher : North-Holland, DOI : [10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- SIMON E., GUIGUE V. & PIWOWARSKI B. (2019). Unsupervised information extraction : Regularizing discriminative approaches with relation distribution losses. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, p. 1378–1387, Florence, Italy : Association for Computational Linguistics. DOI : [10.18653/v1/p19-1133](https://doi.org/10.18653/v1/p19-1133).
- THORNDIKE R. L. (1953). Who belongs in the family ? *Psychometrika*, **18**(4), 267–276. Publisher : Springer, DOI : [10.1007/BF02289263](https://doi.org/10.1007/BF02289263).
- TRAN T. T., LE P. & ANANIADOU S. (2020). Revisiting Unsupervised Relation Extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, p. 7498–7505, Online : Association for Computational Linguistics. DOI : [10.18653/v1/2020.acl-main.669](https://doi.org/10.18653/v1/2020.acl-main.669).