Internship: Span and Arc Representations for Relation Extraction

Joseph Le Roux

November 28, 2024

1 Context

Floquet et al. [Flo+23] have recently proposed a model to represent the units manipulated by combinatorial problems for NLP parsing with vectors. In other words this means that these models compute vectors corresponding to spans for constituent parsing and to arcs for dependency parsing. These models have been shown to reach state-of-the-art performance on various corpora.

Vectors for units are computed from word embeddings and are then used to perform both parsing tasks: structure prediction and labelling. This helps sharing parameters between tasks and has a beneficial impact on parse quality. These vectors can be further refined with Transformers, so unit vectors can incorporate information from other units, thus realizing a soft version of higher-order parsing.

2 Internship Description

For this internship we propose to apply this model to other NLP tasks. In particular we want to focus on joint Named Entity Recognition and Relation Extraction (NER-RE), which can be represented by a model with both spans and arcs [Zar+24]. Spans are used to represent entities and arcs represent relations between entities. The model needs to represent both types of information, and we want to explore scoring models based on approximate MRFs such as [LYT23] Software extension for NER-RE will be based on the code base we already developed for parsing.

3 Application

We are looking for a candidate with NLP background (master level) with good knowledge of Machine Learning methods for NLP. We expect proficiency with python and deep learning libraries such as pytorch. Knowledge of parsing, relation extraction and more generally graph-based approaches to NLP tasks is a plus.

The internship work will be carried out at LIPN at Université Sorbonne Paris Nord, on site (no remote), with possibilities to be extended to a three-year Ph.D. funding (2025-2028)

For additional information, please contact leroux@lipn.fr. If you are interested please attach to your application email a CV, a cover letter and a transcript of your Master level marks.

References

[Flo+23] Nicolas Floquet, Joseph Le Roux, Nadi Tomeh, and Thierry Charnois. “Attention sur les spans pour l’analyse syntaxique en constituants”. In: *18e Conférence en Recherche d’Information et Applications – 16e Rencontres Jeunes Chercheurs en RI – 30e Conférence sur le Traitement Automatique des Langues Naturelles – 25e Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues*. Ed. by Christophe Servan and Anne Vilnat. Paris, France: ATALA, 2023, pp. 37–45. URL: https://hal.science/hal-04130207.

[LYT23] Wei Liu, Songlin Yang, and Kewei Tu. “Structured Mean-Field Variational Inference for Higher-Order Span-Based Semantic Role Labeling”. In: *Findings of the Association for Computational Linguistics: ACL 2023*. Ed. by Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki. Toronto, Canada: Association for Computational Linguistics, July 2023, pp. 918–931. DOI: 10.18653/v1/2023.findings-acl.58. URL: https://aclanthology.org/2023.findings-acl.58.

[Zar+24] Urchade Zaratiana, Nadi Tomeh, Pierre Holat, and Thierry Charnois. “An Autoregressive Text-to-Graph Framework for Joint Entity and Relation Extraction”. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 38.17 (Mar. 2024), pp. 19477–19487. DOI: 10.1609/aaai.v38i17.29919. URL: https://ojs.aaai.org/index.php/AAAI/article/view/29919.