Université de Lorraine - IDMC
Masters SC & TAL – M1
2020-2021

Fiche de projet tutoré / Project form

Évaluer l'accord inter-analyseurs / Evaluate inter-parser agreement

Encadrement / Supervisors

1. SyNaLP team, LORIA
1. encadrant•e principal•e / main supervisor :
Yannick Parmentier (yannick.parmentier@loria.fr)
2. autres encadrant•es / other supervisors : /

Description / Description

1. projet global/global project

Being able to parse a sentence to compute a representation of its internal structure, either in terms of syntactic dependencies or constituents, or even semantic representations (e.g. Abstract Meaning Representations [Banarescu et al., 2013]) is a pre-requesite for many NLP applications, such as text-to-text generation, dialog systems or machine translation.

Over the last decades, many parsers have been developed and trained over large annotated corpuses, resulting in robust parsing environments (see e.g. [Flanigan et al., 2014], [Ji et al., 2019], [Zhang, 2020]).

The efficiency of these parsers has been well studied on goldstandards, using various metrics such as Labelled/Unlabelled Attachment Scores for dependency parsing, Precision / Recall for constituency parsing or semantic parsing. A first point to be noted is that these metrics are useful e.g. to compare between parsers, but do not consider contextual information about e.g. the aim of the parser in a given bigger application. Another important point is that best results are achieved in-domain (parser trained and run on data belonging to the same domain). For many applications targetting out-of-domain data, it may be difficult to assess a given parser's efficiency.

A common way to determine whether the output of a given parser is reliable on a given out-of-domain dataset, is actually to run several parsers on the same data and check their agreement. When parsers have been trained on datasets using the same type of information (e.g. same dependency tagsets), this agreement may be relatively easy to assess. When it is not the case, some alignment is needed. This alignment may be done e.g. by applying transformation rules between pairs of tags.

In this project, we plan to (i) choose an application field (e.g. dependency parsing), (ii) take three freely available parsers among the most efficient ones (see e.g. [Ruder, 2020]), and evaluate how good they perform and agree on out-of-domain data. From this investigation, we will design alignment techniques (transformation-based or not) and evaluate their impact on parsing out-of-domain data (how well does the inter-parser agreement increase match successful parses?).

2. biblio. UE 705 (semestre 7)
The first step of the project consists in gathering a set of bibliographical references related to cross-domain parsing (e.g. [Fried et al., 2019] for constituency parsing). Note that we will not only be interested in articles about domain adaptation, but also about parsing evaluation metrics, such as [Goodman, 1996].
3. réalisation. UE 805 (semestre 8)
In a second step, students are expected to design and implement a inter-parser agreement evaluation experiment, from which alignments techniques would be defined and then included in the experimental setup for validation. Examples of free available parsing environments include the Biaffine Parser [Dozat and Manning, 2017] (see https://github.com/yzhangcs/parser).

Informations diverses : matériel nécessaire, contexte de réalisation /
Various information: material, context of realization
This project is related to the GramEx plateform, whose aim is to parse free text to generate exercises for natural language learners. Being able to assess a parser's efficiency in a given context is a key feature.
Parsers will be used without any specific training. That is, their default training (e.g. on the Universal Dependency treebanks) will be used. Out-of-domain data will come from online open sources (e.g. wikipedia).

Livrables et échéancier / Deliverable and schedule
- November-December: bibliography
- February: design of experimental setup (selection of parsers and evaluation metrics)
- March-April: implementation and testing
- May: analysis of the results, report writing and defense

Bibliographie /References (max. 4-5)

[Dridan and Oepen, 2013] Rebecca Dridan and Stephan Oepen: "Document parsing: Towards realistic syntactic analysis" in *Proceedings of the 13th International Conference on Parsing Technologies*, Nara, Japan (2013).

[Goodman, 1996] Goodman, Joshua : «Parsing Algorithms and Metrics» in *Proceedings of the 34th Meeting of the Association for Computational Linguistics*, Santa Cruz, USA (1996).

[Fried et al., 2019] Daniel Fried, Nikita Kitaev and Dan Klein : «Cross-Domain Generalization of Neural Constituency Parsers» in *Proceedings of the 57th Conference of the Association for Computational Linguistics*, ACL 2019, Florence, Italy (2019), pp. 323-330.

[Ruder, 2020] Sebastian Ruder : «NLPProgress» http://nlpprogress.com/

[Smith and Eisner, 200] David A. Smith, Jason Eisner : «Parser Adaptation and Projection with Quasi-Synchronous Grammar Features», in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, *Singapore* (2009).
.
[Tsarfaty et al., 2011] Reut Tsarfaty, Joakim Nivre, Evelina Andersson : «Evaluating dependency parsing: robust and heuristics-free cross-annotation evaluation», in *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (2011), pp 385–396.