

2022-05571 - Evaluation of the Machine Translation of Scientific Documents

Type de contrat : Stage
Niveau de diplôme exigé : Bac + 4 ou équivalent
Fonction : Stagiaire de la recherche

Contexte et atouts du poste

This internship will take place in the context of the ANR project MaTOS (Machine Translation for Open Science), which aims to develop new methods of automatically translating and evaluating scientific documents. The project focuses on translation between English and French, for which resources are readily available and translations are of a reasonable quality and coherence. The internship could potentially lead to a PhD thesis starting in September 2023 financed by MaTOS. The internship will be supervised by Rachel Bawden and will involve collaborations with the other partners in the project, notably François Yvon (CNRS).

The length of the internship is 6 months starting on the 1st March 2023 at the earliest.

Mission confiée

The topic of this internship is the evaluation of machine translation (MT) of scientific documents. The automatic evaluation of MT is a crucial component of model development and remains a challenging subject. The development of automatic metrics, which seek to replicate human judgments of translation quality, is a major area of study, and many metrics exist, from simple ones that rely on counting lexical overlap such as BLEU (Papineni et al., 2002) and METEOR (Banerjee and Lavie, 2005) to those relying on more recent techniques (e.g. pre-trained neural language models) such as BERTScore (Zhang et al., 2020), BARTScore (Yuan et al., 2021), BLEURT (Sellam et al., 2020) and COMET (Rei et al., 2020). Besides the general challenges faced when defining MT metrics, the evaluation of the MT of scientific documents poses specific challenges, one of them being the heavy use of domain-specific terms, which, if translated incorrectly, severely impact the quality of the translation. Evaluation metrics should therefore also be sensitive to specific challenges faced by the evaluation of scientific documents: (i) the correct translation of terms, (ii) the coherent translation of terms within a document (with respect to term variants, use of acronyms, etc.) and (iii) the capacity to maintain a logical argument between sentences and sections. Previous work has suggested provided complementary measures to evaluate these specific aspects such as correct term translation (Alam et al., 2021) and lexical cohesion (Wong and Kit, 2012).

This internship will involve exploring alternative ways of evaluating terminological aspects of scientific document translation. Inspired by the use of question-based metrics to evaluate text generation tasks (Scialom et al., 2021), one possible direction is to explore how terminologies, relation extraction and information extraction can be used as a means of evaluation of translation quality. For example, (i) can the same relations be found between a reference (human-produced) translation and an automatically produced one? (ii) can terms be matched in similar parts of the document? (iii) how coherent is the use of terms within a document? and (iv) can the same information be extracted from an MT output and the source or reference text? The internship will involve both analysis of existing data, for instance the biomedical translation task (Bawden et al., 2020; Yeganova et al., 2021) and the use and training of neural NLP models.

References

Md Mahfuz Ibn Alam, Antonios Anastasopoulos, Laurent Besacier, James Cross, Matthias Gallé, Philipp Koehn, and Vassilina Nikoulina. On the evaluation of machine translation for terminology consistency, June 2021.

Satanjeev Banerjee and Alon Lavie. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, pages 65–72, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics. <https://aclanthology.org/W05-0909>.

Rachel Bawden, Giorgio Maria Di Nunzio, Cristian Grozea, Inigo Jauregi Unanue, Antonio Jimeno Yepes, Nancy Mah, David Martínez, Aurélie Névél, Mariana Neves, Maite Oronoz, Olatz Perez-de Viñaspre, Massimo Piccardi, Roland Roller, Amy Siu, Philippe Thomas, Federica Vezzani, Maika Vicente Navarro, Dina Wiemann, and Lana Yeganova. Findings of the WMT 2020 biomedical translation shared task: Basque, Italian and Russian as new additional languages. In Proceedings of the Fifth Conference on Machine Translation, pages 660–687, Online, November 2020. Association for Computational Linguistics. <https://aclanthology.org/2020.wmt-1.76>.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pages 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics. doi: 10.3115/1073083.1073135. <https://aclanthology.org/P02-1040>.

Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. COMET: A neural framework for MT evaluation. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 2685–2702, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.213. <https://aclanthology.org/2020.emnlp-main.213>.

Thomas Scialom, Paul-Alexis Dray, Sylvain Lamprier, Benjamin Piwowarski, Jacopo Staiano, Alex Wang, and Patrick Gallinari. QuestEval: Summarization asks for fact-based evaluation. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 6594–6604, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-main.529. <https://aclanthology.org/2021.emnlp-main.529>.

Thibault Sellam, Dipanjan Das, and Ankur Parikh. BLEURT: Learning robust metrics for text generation. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 7881–7892, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.704. <https://aclanthology.org/2020.acl-main.704>.

Billy T. M. Wong and Chunyu Kit. Extending machine translation evaluation metrics with lexical

Informations générales

- **Thème/Domaine** : Langue, parole et audio Systèmes d'information (BAP E)
- **Ville** : Paris
- **Centre Inria** : Centre Inria de Paris
- **Date de prise de fonction souhaitée** : 2023-03-01
- **Durée de contrat** : 6 mois
- **Date limite pour postuler** : 2022-12-31

Contacts

- **Equipe Inria** : ALMANACH
- **Recruteur** :
Bawden Rachel / rachel.bawden@inria.fr

A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 200 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3500 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 180 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

L'essentiel pour réussir

We especially welcome candidates who already have experience in NLP and who have a high-level of motivation for the topic and the task, as well as the capacity to propose new ideas.

Consignes pour postuler

Sécurité défense :

Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.

Attention: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

cohesion to document level. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pages 1060–1068, Jeju Island, Korea, July 2012. Association for Computational Linguistics. <https://aclanthology.org/D12-1097>.

Lana Yeganova, Dina Wiemann, Mariana Neves, Federica Vezzani, Amy Siu, Inigo Jauregi Unanue, Maite Oronoz, Nancy Mah, Aurélie Névéol, David Martinez, Rachel Bawden, Giorgio Maria Di Nunzio, Roland Roller, Philippe Thomas, Cristian Grozea, Olatz Perez-de Viñaspre, Maika Vicente Navarro, and Antonio Jimeno Yepes. Findings of the WMT 2021 biomedical translation shared task: Summaries of animal experiments as new test set. In Proceedings of the Sixth Conference on Machine Translation, pages 664–683, Online, November 2021. Association for Computational Linguistics. <https://aclanthology.org/2021.wmt-1.70>.

Weizhe Yuan, Graham Neubig, and Pengfei Liu. BARTScore: Evaluating generated text as text generation. In Curran Associates, Inc., editor, Advances in Neural Information Processing Systems, pages 27263–27277, Online, 2021.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. BERTScore: Evaluating text generation with BERT. In 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26–30, 2020. OpenReview.net, 2020. <https://openreview.net/forum?id=5keHuCVFDr>.

Principales activités

The main activities will be to carry out research on the topic outlined by (i) studying the past literature, (ii) re-implementing previously proposed approaches and baselines, (iii) proposing improvements to these solutions or a novel approach, (iv) carrying out and writing up experiments, (v) communicating on those experiments with the group.

Compétences

Candidates should be currently finishing a Master 2 or equivalent (e.g. engineering school) in computer science (speciality artificial intelligence, machine learning or natural language processing).

They should have a good level in programming (python), experience with neural networks and an interest in natural language processing. A good written and spoken level of English is required, and knowledge of French is preferred.

Avantages

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Rémunération

This internship may be either compensated (3,90€/hour) or remunerated (SMIC = 1 678,95€) depending on the candidate's situation.