

第11回アジア翻訳ワークショップ (WAT2024)開催報告

Report of the 11th Workshop on Asian Translation (WAT2024)



東京大学 大学院情報理工学系研究科 特任研究員

中澤 敏明

2010年京都大学大学院情報学系研究科知能情報学専攻博士課程修了。博士（情報学）。現在は東京大学大学院情報理工学系研究科特任研究員。機械翻訳の研究に従事。

✉ nakazawa@nlab.ci.i.u-tokyo.ac.jp

① はじめに

アジア翻訳ワークショップ（Workshop on Asian Translation, WAT）はアジア言語を中心とした評価型機械翻訳ワークショップであり、2014年に第1回（WAT2014）を開催して以降、毎年開催している。本稿の著者は初回からオーガナイザーの一人としてワークショップの運営を行っている。2016年の第3回（WAT2016）以降は自然言語処理の国際会議との併設ワークショップとして開催しており、2024年の第11回（WAT2024[1]）はアメリカのマイアミで11月12日から16日に開催されたEMNLP2024の併設ワークショップとして開催された。参加者は現地参加とオンラインと合わせて約20名程度で、多くはオンラインでの参加であった。

② 研究論文

WAT2024では5件の研究論文を採択した。採択した研究論文のタイトルを以下に示す。

- Machine Translation Of Marathi Dialects: A Case Study Of Kadodi
- AI-Tutor: Interactive Learning of Ancient Knowledge from Low-Resource Languages
- An Empirical Study of Multilingual Vocabulary for Neural Machine Translation Models
- Are Large Language Models State-of-the-art Quality Estimators for Machine Translation of

User-generated Content?

- Creative and Context-Aware Translation of East Asian Idioms with GPT-4

この5件の中で、査読の評価が最も高かった“Machine Translation Of Marathi Dialects: A Case Study Of Kadodi”をbest paperとして選出した。

③ パネルディスカッション

WAT2024では招待講演のかわりにパネルディスカッションを行なった。パネリストとして蘇州大学のMin Zhang氏（ビザの関係で残念ながら当日出席できず）、National Electronics and Computer Technology Center (NECTEC)のThepchai Supnithi氏、川村インターナショナルの森口功造氏、TransPerfectのFred Bane氏をお迎えし、以下のようなテーマで討論を行なった。

- Current status and future of MT in your country
- How is MT adopted by human translators in your country
- What should the MT researchers do for practical MT
- NMT vs. LLM MT for human translators
- Future of MT
- etc.

パネルディスカッションの中では「NMT/LLMの技術の利用はどんどん広がっているが、適切に利用しない

と良い結果は得られない」「NMTの長所は一貫した出力が出せること、ドメイン適応に優れていること」「LLMの長所はプロンプトによる高いカスタマイズ性、長い文脈の理解、より流暢な翻訳」「LLMの短所は速度が遅い、費用が高い、英語以外の言語における精度が低い」「今後必要なことは多言語化、マルチモーダル、データ不足の解消、セキュリティーなど」といった意見が出た。また、タイではそもそも自然言語処理の研究者が少ないといった問題点も挙げられた。

また、今後も定期的にこのような情報交換の場を設け、アジア全体での機械翻訳コミュニティの活性化を目指し、WATがそのプラットフォームとしての役割を果たすべきであるという結論となった。

ワークショップ全体の録画はEMNLPに参加登録を行った方であれば、以下のURLから視聴できる。

<https://underline.io/events/470/sessions?eventId=18628&searchGroup=lecture&searchGroup=lecture>

4 翻訳タスク

WAT2024は世界最大の機械翻訳のワークショップであるWMT2024と同じ会場で開催していた。そのため参加者の利便性を図るため、今回に限り、翻訳タスクはWMTの翻訳タスクとして実施した。実施したタスクは以下の4つである。

- Multilingual22MT [2]
- English-to-Lowres Multi-Modal Translation [3]
- Non-Repetitive Translation [4]
- Patent Translation [5]

どのタスクにも数チームの参加があった。それぞれのタスクの結果は、それぞれのFindings論文にまとめられているので、興味のある方は参考にしていきたい。

5 まとめ

本稿ではWAT2024全体の概要を報告した。残念ながらWAT2024では当日の参加者が少なかったが、論文の投稿数や翻訳タスクの参加者は一定数あり、またパネルディスカッションも有意義であったことから、ワークショップとしては成功だったと言える。

WATは今後も継続して開催予定である。2025年度は12月にインドのムンバイで開催されるAAACL-IJCNLP2025のワークショップとしての開催が決定している。WAT2025では新たに日英の特許翻訳、特に請求項に特化した翻訳のタスクを実施予定である。請求項は1文が長い、前の文脈を考慮して翻訳する必要がある、書き方のスタイルが独特であるなどの特徴があり、特許翻訳の中でもかなり翻訳が難しい部分である。そのため非常にチャレンジングなタスクであると言える。このタスクでは翻訳精度の向上ももちろんだが、評価の部分に力を入れる予定である。WMTでも近年よく言われていることだが、近年主流であるベクトルベースの評価手法は、ドメインによっては正確な評価を行うことができない。おそらく特許翻訳に関しても同じ議論ができるはずで、特許翻訳ドメインにおいてはいまだにBLEU評価の方が信頼性が高い可能性もある。これらの点について明らかにし、特許翻訳用の評価基準を作ることが本タスクの大きな目標である。

なおWATでは評価を行うための費用等のためのスポンサーを募集しているため、興味のある方はご連絡いただければ幸いである。

参考文献

- [1] Toshiaki Nakazawa and Isao Goto. 2024. Proceedings of the Eleventh Workshop on Asian Translation (WAT 2024). Association for Computational Linguistics, Miami, Florida, USA.
- [2] Raj Dabre and Anoop Kunchukuttan. 2024. Findings of WMT 2024's Multilingual22MT Shared Task for Machine Translation of 22 Indian Languages. In Proceedings of the Ninth Conference on Machine Translation, pages 669-676, Miami, Florida, USA. Association for Computational Linguistics.
- [3] Shantipriya Parida, Ondřej Bojar, Idris Abdulmumin, Shamsuddeen Hassan Muhammad, and Ibrahim Said Ahmad. 2024. Findings of WMT2024 English-to-Low Resource Multimodal Translation Task. In Proceedings of the Ninth Conference



on Machine Translation, pages 677–683, Miami, Florida, USA. Association for Computational Linguistics.

[4] Kazutaka Kinugawa, Hideya Mino, Isao Goto, and Naoto Shirai. 2024. Findings of the WMT 2024 Shared Task on Non-Repetitive Translation. In Proceedings of the Ninth Conference on Machine Translation, pages 715–727, Miami, Florida, USA. Association for Computational Linguistics.

[5] Shohei Higashiyama. 2024. Results of the WAT/WMT 2024 Shared Task on Patent Translation. In Proceedings of the Ninth Conference on Machine Translation, pages 118–123, Miami, Florida, USA. Association for Computational Linguistics.



4
1

機械翻訳