

MutNMT, an open-source NMT tool for educational purposes

Gema Ramírez Sánchez
Prompsit Language Engineering, S.L.
gramirez@prompsit.com

We present MutNMT,¹ an open-source web application for educational purposes to introduce non-experts to NMT. The tool, developed within the MultiTraiNMT project² along with other training materials (a book³ and activities⁴), gathers the feedback of academic and industrial project partners and also external collaborators.

MutNMT is based on open-source code from JoeyNMT (Kreutzer et al., 2019),⁵ an open-source minimalist neural machine translation toolkit also for educational purposes. It uses its Transformer architecture to train NMT models. A new feature to extract the n-best list of translation candidates was contributed to JoeyNMT from MutNMT.

MutNMT provides a user-friendly interface to manage the full process of building an NMT system, provided that training data is available, through different sections: 1) data set uploader and library for own and shared data sets, 2) engines library providing access to own or shared models, their details and a full training log, 3) training section where users can select corpora and set training parameters, 4) translation section for short texts and documents, 5) model inspection and comparison among models section, 6) evaluation with automatic metrics and, finally 7) administration of users and monitoring of processes and server.

The tool was conceived for an educational environment with very limited computational capabilities, e.g. a server with 2-4 GPUs. There are

3 different roles: beginners (not allowed to train), experts (allowed to train) and administrators. To train a new model, 1-hour training slots are allocated and can be resumed for an additional hour slot. To provide a good experience to users, optimal ranges for data sizes and training parameters were estimated. Training sets are limited to 500k sentence pairs, and development and test sets to 5k. Customisable parameters as vocabulary, beam or batch size, validation frequency, stopping condition and duration are also constrained.

An engaged community of users is starting to arise around MutNMT. Some instances have been deployed successfully in different universities and are being actively used to train students and professional translators. The MultiTraiNMT project official instance provided by the Universitat Autònoma de Barcelona⁶ currently has 560 users from a variety of companies and research institutions mainly related to translation technology training. Access only requires a Google account or no requirement for the demo version. Further development plans include new roles for research-oriented experiments with more flexible training time and data sizes, API usage, usage integrated with CAT-tools and further evaluation metrics.

References

Kreutzer, Julia, Jasmijn Bastings, and Stefan Riezler. 2019. Joey NMT: A minimalist NMT toolkit for novices. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*, pages 109–114, Hong Kong, China, November. Association for Computational Linguistics.

© 2023 The authors. This article is licensed under a Creative Commons 4.0 licence, no derivative works, attribution, CC-BY-ND.

¹<https://github.com/Prompsit/mutnmt>

²<https://www.multitrainmt.eu/index.php/>

³Machine translation for everyone: Empowering users in the age of artificial intelligence <https://langsci-press.org/catalog/book/342>

⁴<https://github.com/jaspock/mt4everyone>

⁵<https://github.com/joeynmt/joeynmt>

⁶<https://ntradumatica.uab.cat/>