Dynamic Adaptation of Neural Machine-Translation Systems Through Translation Exemplars

Arda Tezcan
LT[3], Language and Translation Technology Team
Ghent University
Belgium
Arda.Tezcan@ugent.be

Abstract

This project aims to study the impact of adapting neural machine translation (NMT) systems through similar translations retrieved from translation memories, determine the optimal metric(s) for measuring similarity, and, verify the usefulness of this approach for domain adaptation of NMT systems.

1 Introduction

Translation exemplars, i.e. previously observed or generated translations, whose source is similar to new input, play an important role in the translation process by providing explicit information on context, exceptions, and irregularities, which are difficult to generalise. In the context of machine translation (MT), the information provided by similar translations can also be considered complementary to the neural models, which are good at making generalisations. "Dynamic Adaptation of Neural Machine-Translation Systems Through Translation Exemplars" is a three-year, post-doctoral research project (October 2020 – September 2023), funded by the Research Foundation – Flanders (FWO) with the following scientific objectives: (i) study the impact of adapting NMT systems through similar translations retrieved from TMs, (ii) determine the optimal metrics for measuring translation similarity, and (iii) verify the usefulness of this approach for domain adaptation. All research activities are carried out at the Department of Translation, Interpreting and Communication, Ghent University, Belgium.

The project has been built upon the methodology used in the translation memory (TM) — NMT integration approach, neural fuzzy repair (NFR) (Bulté and Tezcan, 2019).[1] In NFR, each source sentence in the TM is augmented (concatenated), with the translation of the most similar fuzzy match (FM), retrieved from the same TM using edit distance, when at least one FM is found above the minimum similarity threshold of λ. The augmented TM is merged with the original TM to train an NMT model. During inference, the same technique is applied to augment source sentences prior to obtaining translations from the trained model. If no FMs above the minimum similarity threshold can be found, the translations are obtained for the original source sentence, using the same model.

Tests on multiple language pairs showed that this method results in substantial gains in translation quality compared to (i) baseline MT systems, (ii) the FMs themselves when used as final output, even when they correspond to near-perfect translations in high similarity ranges, and (iii) a 'fuzzy match repair' approach, which relies on editing highly similar FMs to arrive at the final translation (Bulté et al., 2018). These experiments also demonstrated that using the NFR system starts being advantageous with a minimum FM score of $\lambda = 0.5$, and augmenting source sentences with FMs of higher similarity scores leads to higher translation quality. While this study showed the usefulness of adapting NMT systems through similar translations, it also led to new research questions and formed the scientific objectives of this project. The following sections provide an overview of the progress made on the first

[1] https://github.com/lt3/nfr

two scientific objectives and summarise the plans for future work.

2 Optimal metrics for measuring translation similarity

Text similarity can be measured from different perspectives, such as using string-, semantic- and syntactic-level information. In the context of NFR, it is yet to be determined if, and to what extent, different similarity levels are important in retrieving informative FMs and producing better translations.

To seek optimal similarity metrics in the context of NFR, string- and semantic-similarity metrics have been studied in combination with different sub-word segmentation methods (Tezcan et al., 2021). Retrieving FMs by measuring cosine similarity between sentence embeddings resulted in translations with higher quality in comparison to using edit distance. Moreover, applying sub-word segmentation prior to measuring semantic similarity improved translation quality further.

To utilise sub-sentence-level similarities between two sentences more explicitly, two additional approaches have been tested: (i) marking relevant tokens in retrieved FMs, and (ii) augmenting source sentences with multiple FMs that lead to a maximum coverage of source tokens (Tezcan et al., 2021). When combined, these methods led to improvements in estimated translation quality for 8 language directions (English $\leftrightarrow$ Dutch, French, Hungarian, Polish), compared to a baseline transformer NMT model and the original NFR approach (Bulté and Tezcan, 2019).

3 The impact of adapting NMT systems through similar translations

To analyse the impact of adapting NMT systems through similar translations on translation quality, evaluations were carried out both automatically and manually.[2]

The fine-grained, human error analysis showed that both the adapted and non-adapted NMT systems made a comparable number of errors. On the other hand, while the adapted NMT system produced more fluent translations, with a significant reduction in *lexicon* and *coherence* errors, it also diverged from the source content and meaning (i.e. reduced accuracy) more often than the baseline NMT system, making more errors of *addition* and *mistranslation*.

The automatic evaluation analysed translation quality by relying on the reference translations while using metrics that targeted lexical, sub-lexical, semantic, and syntactic aspects of translation quality. According to all evaluation metrics, the quality of the adapted NMT system was estimated to be higher than that of those produced by the NMT system. The difference in quality was always significant and large confirming that the improvements were obtained for all the different aspects of quality that were analysed.

By analysing the MT output with different resources (source–MT vs. MT–reference), the two evaluation methods yielded a more nuanced picture of the differences in translation quality between the two systems. They also revealed an interesting property of the NFR system: by using the similar translations, it was able to learn systematic deviations from the source text (e.g. related to the use of cohesive devices or translation decisions affecting sentence boundaries) and produce translations that are similar to the reference translations, even though such deviations were marked as accuracy errors during human error analysis.

4 Future work

Future work will focus on (i) studying the usefulness of syntactic similarity for retrieving informative FMs, (ii) improving the NFR performance further, and (iii) adopting this methodology to the task of domain adaptation.

[2]Evaluations were made using the best-performing NFR methodology as outlined in Section 2.

References

Bulté, Bram and Arda Tezcan. 2019. Neural Fuzzy Repair: Integrating Fuzzy Matches into Neural Machine Translation. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, 1800–1809.

Bulté, Bram and Tom Vanallemeersch and Vincent Vandeghinste. M3TRA: integrating TM and MT for professional translators. *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, Alicante, Spain, 69–78.

Tezcan, Arda and Bram Bulté. 2022. Evaluating the Impact of Integrating Similar Translations into Neural Machine Translation. *Information, 13(1):19.*

Tezcan, Arda, Bram Bulté, and Bram Vanroy. 2021. Towards a Better Integration of Fuzzy Matches in Neural Machine Translation through Data Augmentation. *Informatics-Basel 8(1):7.*