Improving Domain Robustness in Out-of-Domain Corpus

Ethelbert Uzodinma[1], Jennifer Spenader[2], Joost Doornkamp[3]

[1] University of Groningen, 9700 AB Groningen, Netherlands
[2] University of Groningen, 9700 AB Groningen, Netherlands
euetel@gmail.com[1], e.uzodinma@student.rug.nl[1]
j.k.spenader@rug.nl[2] , j.doornkamp@rug.nl[3]

Abstract. Despite recent successes with neural machine translation, even models trained on extensive parallel corpus data still perform poorly when applied to translating text from a different genre, so called "out-of-domain" translation. This thesis investigates different methods that can be used to improve out-of-domain translations such as byte-pair encoding, sub-word regularization, beam size, label smoothing and domain-adaptation. These techniques were applied to the training and fine-tuning stages in multiple out-of-domain translation experiments. The results showed improvement in fluency and adequacy. Evaluations using BLEU scores and perplexity showed an overall improvement of above 10% in out-of-domain translations.

Keywords: Domain robustness · In-domain · Out-of-domain · Machine translation · Hallucination.

1 Introduction

One of the problems that have hindered the advancement of Neural Machine Translation (NMT) is domain shift which can lead to a condition known as "Hallucination" [4],[5]. This condition is identified in NMTs when the translation is grammatically correct (fluency) but has a different semantic meaning from the source text (adequacy).

The area of focus in this research is to overcome the challenges of MT in the area of domain mismatch which leads to "Hallucination". The consequence of this condition is the inability of MT systems failing in performance and not generalizing well when translating text outside the domain of the training set. When this occurs the MT model can be said not to be robust enough, which leads to the concept of Domain Robustness [5],[6].

Domain robustness can be described as the ability of a machine translation model to generalize well to unseen data from other domains. This concept of domain robustness satisfies the main goal of machine translation which is to learn models that generalize well to a data distribution outside the distribution of the training set.

2 Methods

The OPUS dataset was used for the training, validation and testing in a proportion of 80%, 10%, 10% respectively. It consists of 5 genres of parallel corpus of German to English language translations, with domains from the Medical, Law, IT, Koran and Open subtitles. The IT domain was augmented with 5 other corpora within the same genre because of insufficient data. The transformer[3] architecture was used for training and model generation.

The machine translation workflow was implemented in 3 stages, this include: text pre-processing, training and fine-tuning. The text pre-processing stage consists of unicode normalization, punctuation removal, lower casing, tokenization and filtering. During the training stage, different combinations of byte-pair encoding(BPE) algorithm at 32K merge operations, sub-word regularization strategy(i.e BPE dropout) of 10% of each merge operation and 10% label smoothing was applied to the source and target text respectively.

The model was trained from scratch on the Medical(in-domain) dataset and a small percentage of out-of-domain data from any of the other 4 domains using various hyper-parameter settings of adaptive learning rate with momentum (Adam) and cross entropy loss function at 45 epochs etc. The regularization strategies used include: label smoothing, BPE dropout and early stopping to prevent overfitting and over-confidence, while domain adaptation was applied during the fine-tuning stage on the other domains using the same hyper-parameter settings. A brief description about this method is highlighted below, for detailed information on the methods refer to (Uzodinma, 2022[6]).
Domain Adaptation [1]: a type of transfer learning technique applied during the fine-tuning stage of the machine translation. It transforms the parameter space of the model that was pre-trained on the large dataset(Medical) to a new feature space of the small dataset(IT). Other methods that could possibly work include zero or few-shot learning.

3 Evaluation/Results

The model was evaluated based on BLEU scores[2] and perplexity metrics. Then visualized with an attention matrix map in which higher correlation was observed between the source and prediction. Furthermore, a lot of improvement was seen in terms of corpus-level BLEU scores above the baseline. In 5 translation tasks across the domains, there was at least a gain of +1.5 BLEU scores and 20% gain in BLEU scores respectively. The domain adaptation technique showed best performance when compared with the other combination of techniques during the experiments with the out-of-domain data. The overall conclusion from the experiments across domains shows that the law text showed best improvement compared to the text from the other domains, this could be due to a number of reasons which include that the data distribution of the law text is the most related to the medical text while the open-subtitles showed least improvement.

Table 1. Number of hallucinations across the domains at a threshold of 1.0 BLEU, *Where BPE=byte-pair encoding, SWR= sub-word regularization, DA = domain adaptation*

Domain	% hallucinations NMT + BPE	% Hallucinations NMT + BPE +SWR	% Hallucinations NMT + BPE + DA
Law	41	38.1	22.5
IT	59.65	56.15	43.1
Open Subtitles	61	59.25	53.7
Koran	60.7	59.3	52.7
Average	55.6	51.9	41.3

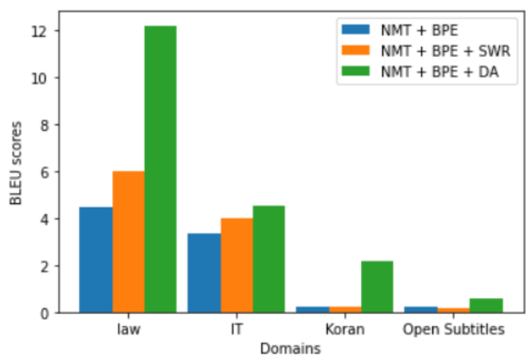

Fig. 1. Summary of BLEU scores across domains.

References

1. Hu, J., Xia, M., Neubig, G., Carbonell, J.: Domain Adaptation of Neural Machine Translation by Lexicon Induction. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy. Association for Computational Linguistics. 2019 (pp 2989–3001).
2. Koehn, P., Monz, C.: Manual and automatic evaluation of machine translation between european languages. InProceedings on the Workshop on Statistical Machine Translation 2006 Jun (pp. 102-121).
3. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit J., Jones L., Gomez AN., Kaiser Ł., Polosukhin I.: Attention is all you need. Advances in neural information processing systems 30. 2017
4. Wang, C., Sennrich, R.: On exposure bias, hallucination and domain shift in neural machine translation. arXiv preprint arXiv:2005.03642. 2020 May 7.
5. Müller, M., Rios, A., and Sennrich, R.: Domain robustness in neural machine translation," In Proceedings of the 14th Conference of the Association for Machine Translation in the Americas. vol. 1 , 2020 Oct. (pp. 151–164)
6. Uzodinma, E.: Improving domain robustness in translating out-of-domain corpus. Msc thesis Artificial Intelligence, Bernoulli institute, university of Groningen. 2022 Aug. 31.