Domain Adaptation for Neural Machine Translation

Danielle Saunders*
Department of Engineering
University of Cambridge
Cambridge, UK
ds636@cantab.ac.uk

The development of deep learning techniques has allowed Neural Machine Translation (NMT) models to become extremely powerful, given sufficient training data and training time. However, such translation models struggle when translating text of a new or unfamiliar domain (Koehn and Knowles, 2017). A domain may be a well-defined topic, text of a specific provenance, text of unknown provenance with an identifiable vocabulary distribution, or language with some other stylometric feature.

NMT models can achieve good translation performance on domain-specific data via simple tuning on a representative training corpus. However, such data-centric approaches have negative side-effects, including over-fitting and brittleness on narrow-distribution samples and catastrophic forgetting of previously seen domains.

This thesis focuses instead on more robust approaches to domain adaptation for NMT. We consider the case where a system is adapted to a specified domain of interest, but may also need to accommodate new language, or domain-mismatched sentences. As well, the thesis highlights that lines of MT research other than performance on traditional 'domains' can be framed as domain adaptation problems. Techniques that are effective for e.g. adapting machine translation to a biomedical domain can also be used when making use of language representations beyond the surface-level, or when encouraging better machine translation of gendered terms.

Over the course of the thesis we pose and answer five research questions:

How effective are data-centric approaches to NMT domain adaptation? We find that simply selecting-domain relevant training data and fine-tuning an existing model achieves strong results, especially when a domain-specific data curriculum is used during training. However, we also demonstrate the side-effects of exposure bias and catastrophic forgetting.

Given an adaptation set, what training schemes improve NMT quality? We investigate two variations on the NMT adaptation algorithm, regularized tuning including Elastic Weighting Consolidation, and a new variant of Minimum Risk Training. We show they can mitigate the pitfalls of data-centric adaptation. Aside from avoiding the failure modes of data-centric methods, we show these methods may also give better model convergence.

Can domain adaptation help when the test domain is unknown? Most approaches to domain adaptation in the literature assume any unseen test data of interest has a known, fixed domain, with a matching set of tuning data. This thesis works towards relaxing these assumptions. We show that adapting sequentially across domains with regularization can achieve good cross-domain performance without knowing the specific test domain. We also explore domain adaptive model ensembling and automatic model selection. We find this can outperform oracle approaches, which select the best model for inference by using known provenance labels.

Can changing data representation have similar effects to changing data domain? Unlike data domain, data representation – for example, choice of subword granularity or use of syntactic annotation – does not change meaning or correspond to provenance. However, like domain, it can affect the information available to the model, and

*Now at RWS Language Weaver

therefore impacts NMT quality for a given input. We combine multiple representations in a single model or in ensembles in a way reminiscent of multi-domain translation. In particular, we develop a scheme for ensembles of models producing multiple target language representations, and show that multi-representation ensembles improve syntax-based NMT.

Can gender bias in NMT systems be mitigated by treating it as a domain? We show that translation of gendered language is strongly influenced by vocabulary distributions in the training data, a hallmark of a domain. We also show that data selection methods have a strong effect on apparent NMT gender bias. We apply techniques from elsewhere in the thesis to tune NMT on a 'gender' domain, specifically regularized adaptation and multi-domain inference. We show this can improve gendered language translation while maintaining generic translation quality.

Human language itself is constantly adapting, and people's interactions with and expectations of MT are likewise evolving. With this thesis we hope to draw attention to the possible benefits and applications of different approaches to adapting machine translation. We hope that future work on adaptive NMT will focus not only on the language of immediate interest but the machine translation abilities or tendencies that we wish to maintain or abandon.

Acknowledgments

The author would like to thank her PhD supervisor, Bill Byrne. The work was supported by EPSRC grants EP/M508007/1 and EP/N509620/1, with some experiments performed using resources from the Cambridge Tier-2 system operated by the University of Cambridge Research Computing Service[1] funded by EPSRC Tier-2 capital grant EP/P020259/1.

References

Koehn, Philipp and Rebecca Knowles. 2017. Six Challenges for Neural Machine Translation *Proceedings of the First Workshop on Neural Machine Translation*, Vancouver, 28–39.

[1] `http://www.hpc.cam.ac.uk`