Sparse Word Embedding for Multi-domain Machine Translation

Anonymous NAACL submission

Abstract

In this paper, we introduce an application of an old technique [Daume III, 2007] in Multi-Domain Neural Machine Translation. By defining general information encoding region and domain-specific encoding region in word embedding vector, we are able to seperate domain-related information flows in forward and backward propagation during training mitigate the catastrophic interference happens between the inference of different domains in Neural Machine Translation system, i.e to improve the performance of the system for every domain in which it was trained.

1 Introduction

Despite fast improvement thanks to new architectures [Vaswani et al., 2017] [Bahdanau et al., 2015] [Kalchbrenner and Blunsom, 2013] [Sutskever et al., 2014], Machine Translation still struggles with data scarcity. Domain adaptation, or it's special case Multi-domain, is the situation where in-domain data for example banking is rare while out-domain data such as news are plenty, one aims to leverage out-domain data to improve quality of Machine Translation system which is perfectible while being trained with small in-domain dataset. The problem is that we have not been able to apply Domain-adaptation to several domains at once, i.e to build one model which is best for several domains. It is obvious to realize that the bias to in-domain dataset makes model worse in translating other domains. This phenomenon can be implied by well-known problem in Machine Learning, Catastrophic Interference [McCloskey and Cohen, 1989]. In Multi-domain Machine Translation, one have to deal with polysemies whose meaning can be totally different in different domains, for example, the word "chair" means an household tool in "if you are taking a medicine which may cause low blood pressure when rising from a chair or bed " (extracted from medical corpora EMEA [Tiedemann, 2009]) while the word "chair" means to preside over a meeting in following sentence "The President of the ECB or , in his absence , the Vice-President , chairs the meetings of the Governing Council , the Executive Board and the General Council of the ECB. " (extracted from European Central Bank corpora [Tiedemann, 2009]). Using the same word embedding for polysemy obstructs the model from differentiating the meaning of a polysemy in one domain from it's other meaning in the other domain.

2 Related Work

Multi-domain Machine Translation has largely interested NLP community by its promising applications in industry and its relation to fundamental problems in theory. Researchers have proposed a large range of techniques from Data centric methods to Model centric methods [Chu and Wang, 2018]; [Chu et al., 2017]. Data centric methods have goal to collect related-domain data from existing in-domain data by using different techniques such as scoring relatedness by Language Model ([Moore and Lewis, 2010]; [Axelrod et al., 2011]; [Duh et al., 2013] or by metric on the space of sentence embedding [Wang et al., 2017a] or generating pseudo parallel data [Utiyama and Isahara, 2003]; [Wang et al., 2016]; [Wang et al., 2014]. On the other hand, Model centric approaches focus on NMT models that are specialized for domain adaptation. The novelty can be either traininig objective; for example [Luong and Manning, 2015]; [Sennrich et al., 2016]; [Wang et al., 2017b]; [Chen et al., 2017]; [Miceli Barone et al., 2017]; [Zhang and Xiong, 2018] or architecture [KOBUS et al., 2017]; [Gülçehre et al., 2015]; [Britz et al., 2017], [Zhang et al., 2017]; [Vilar, 2018]; [Thompson et al., 2018]; [Zhang et al., 2016]; [Michel and Neubig, 2018] or the decoding algorithm [Gülçehre et al., 2015]; [Khayrallah et al., 2017]. Beside these works, we could also consider "out-of-domain" works such as sequence labeling tasks [Daume III, 2007]; learning multiple visual domains [Rebuffi et al., 2017]. The problem is recently investigated by several interesting works presented in EMNLP 2018 such as [Platanios et al., 2018]; [Zeng et al., 2018]. [Zeng et al., 2018] create in the encoder Domain-specific gate g_i^r and Domain-shared gate g_i^s which are generated from domain-specific and domain-shared semantic representations of source sentence $E_r(x)$ and $E_s(x)$ respectively and which select information from units of hidden states h_i by elementwise product $h_i^r = g_i^r \odot h_i$; $h_i^s = g_i^s \odot h_i$. h_i^r and h_i^s will be fed to Domain-specific and Domain-shared attentional mechanisms respectively. While introducing new features in the architecture, [Zeng et al., 2018] introduces also new objective which is sum of word-level weighted MT objective and objectives of Domain-classifier in source side and target side and Adversarial Domain-classifier in source side. The author have very good approach to well seperate Domain-shared information flow and Domain-specify information flow and determine their contributions to the inference that mitigates the catastrophic interference happens in the network during forward step.

3 The Model

There is not any changes in the architecture of Neural Machine Translation network except the construction of word embedding.

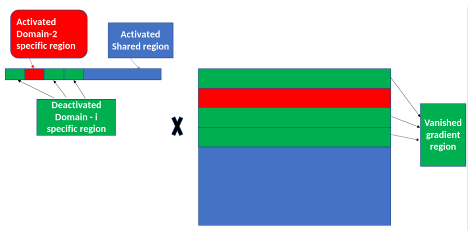

Figure 1: Illustration of the model.

4 Experiments

4.1 Corpora

To evaluate the performance of proposed architecture, we use 4 corpora for training: European Medicines Agency; European Central Bank, News-commentary and European Parliament [Tiedemann, 2009] corresponding to 4 domains: medical, banking, news and administrative respectively and one non-domain corpora Common-crawl which was introduced in Shared Task: Machine Translation of News the Third Conference On Machine Translation 2018(WMT2018). We use Khresmoi-dev, a test set used in the Biomedical Translation Task (WMT2018), newstest2008, and test2006 as validation set for domains: medical, news administrative; except for banking domain, we use a subset of size 2000 extracted from original corpora and excluded from training set. To test the model, we use Khresmoi-test; newstest2009 and test2007 for domains: medical, news administrative; except for banking domain, we use a subset of size 2000 extracted from original corpora and excluded from training set. The information of used corpora is presented in table 1

Task	Corpora	Train	Dev	Test	Vocabulary
English → French	European Medicines Agency	1092568	500	1000	
	European Central Bank	191960	2000	2000	
	News-commentary	258432	2051	3027	
	European Parliament	2007723	2000	2000	
	Common-crawl	3244152			
English → German	European Medicines Agency	1108752	500	1000	
	European Central Bank	109174	2000	2000	
	News-commentary	270769	2051	2525	
	European Parliament	1920209	2000	2000	
	Common-crawl	2399123			

Table 1: Corpora

4.2 NMT engine

4.3 Results

5 Conclusions

Acknowledgments

References

Amittai Axelrod, Xiaodong He, and Jianfeng Gao. 2011. Domain adaptation via pseudo in-domain data selection. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 355–362. Association for Computational Linguistics.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *Proc. ICLR*.

Denny Britz, Quoc Le, and Reid Pryzant. 2017. Effective domain mixing for neural machine translation. In *Proceedings of the Second Conference on Machine Translation*, pages 118–126. Association for Computational Linguistics.

Boxing Chen, Colin Cherry, George Foster, and Samuel Larkin. 2017. Cost weighting for neural machine translation domain adaptation. In *Proceedings* *of the First Workshop on Neural Machine Translation*, pages 40–46. Association for Computational Linguistics.

Chenhui Chu, Raj Dabre, and Sadao Kurohashi. 2017. An empirical comparison of domain adaptation methods for neural machine translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 385–391. Association for Computational Linguistics.

Chenhui Chu and Rui Wang. 2018. A survey of domain adaptation for neural machine translation. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1304–1319. Association for Computational Linguistics.

Hal Daume III. 2007. Frustratingly easy domain adaptation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 256–263. Association for Computational Linguistics.

Kevin Duh, Graham Neubig, Katsuhito Sudoh, and Hajime Tsukada. 2013. Adaptation data selection using neural language models: Experiments in machine translation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 678–683. Association for Computational Linguistics.

Çalar Gülçehre, Orhan Firat, Kelvin Xu, Kyunghyun Cho, Loïc Barrault, Huei-Chi Lin, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2015. On using monolingual corpora in neural machine translation. *arXiv e-prints*, abs/1503.03535.

Nal Kalchbrenner and Phil Blunsom. 2013. Recurrent continuous translation models. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1700–1709. Association for Computational Linguistics.

Huda Khayrallah, Gaurav Kumar, Kevin Duh, Matt Post, and Philipp Koehn. 2017. Neural lattice search for domain adaptation in machine translation. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 20–25. Asian Federation of Natural Language Processing.

Catherine KOBUS, Josep Crego, and Jean Senellart. 2017. Domain control for neural machine translation. In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*, pages 372–378. INCOMA Ltd.

Minh-Thang Luong and Christopher D. Manning. 2015. Stanford neural machine translation systems for spoken language domain. In *International Workshop on Spoken Language Translation*, Da Nang, Vietnam.

Michael McCloskey and Neal J. Cohen. 1989. Catastrophic interference in connectionist networks: The sequential learning problem. *Psychology of Learning and Motivation - Advances in Research and Theory*, 24(C):109–165.

Antonio Valerio Miceli Barone, Barry Haddow, Ulrich Germann, and Rico Sennrich. 2017. Regularization techniques for fine-tuning in neural machine translation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1489–1494. Association for Computational Linguistics.

Paul Michel and Graham Neubig. 2018. Extreme adaptation for personalized neural machine translation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 312–318. Association for Computational Linguistics.

Robert C. Moore and William Lewis. 2010. Intelligent selection of language model training data. In *Proceedings of the ACL 2010 Conference Short Papers*, pages 220–224. Association for Computational Linguistics.

Emmanouil Antonios Platanios, Mrinmaya Sachan, Graham Neubig, and Tom Mitchell. 2018. Contextual parameter generation for universal neural machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 425–435. Association for Computational Linguistics.

Sylvestre-Alvise Rebuffi, Hakan Bilen, and Andrea Vedaldi. 2017. Learning multiple visual domains with residual adapters. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 506–516. Curran Associates, Inc.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Improving neural machine translation models with monolingual data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 86–96. Association for Computational Linguistics.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, pages 3104–3112, Cambridge, MA, USA. MIT Press.

Brian Thompson, Huda Khayrallah, Antonios Anastasopoulos, Arya D. McCarthy, Kevin Duh, Rebecca Marvin, Paul McNamee, Jeremy Gwinnup, Tim Anderson, and Philipp Koehn. 2018. Freezing subnetworks to analyze domain adaptation in neural machine translation. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 124–132. Association for Computational Linguistics.

Jörg Tiedemann. 2009. News from OPUS - A collection of multilingual parallel corpora with tools and interfaces. In N. Nicolov, K. Bontcheva, G. Angelova, and R. Mitkov, editors, *Recent Advances in Natural Language Processing*, volume V, pages 237–248. John Benjamins, Amsterdam/Philadelphia, Borovets, Bulgaria.

Masao Utiyama and Hitoshi Isahara. 2003. Reliable measures for aligning japanese-english news articles and sentences. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc.

David Vilar. 2018. Learning hidden unit contribution for adapting neural machine translation models. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 500–505. Association for Computational Linguistics.

Rui Wang, Andrew Finch, Masao Utiyama, and Eiichiro Sumita. 2017a. Sentence embedding for neural machine translation domain adaptation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 560–566. Association for Computational Linguistics.

Rui Wang, Masao Utiyama, Lemao Liu, Kehai Chen, and Eiichiro Sumita. 2017b. Instance weighting for neural machine translation domain adaptation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1482–1488. Association for Computational Linguistics.

Rui Wang, Hai Zhao, Bao-Liang Lu, Masao Utiyama, and Eiichiro Sumita. 2014. Neural network based bilingual language model growing for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 189–195. Association for Computational Linguistics.

Rui Wang, Hai Zhao, Bao-Liang Lu, Masao Utiyama, and Eiichiro Sumita. 2016. Connecting phrase based statistical machine translation adaptation. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3135–3145. The COLING 2016 Organizing Committee.

Jiali Zeng, Jinsong Su, Huating Wen, Yang Liu, Jun Xie, Yongjing Yin, and Jianqiang Zhao. 2018. Multi-domain neural machine translation with word-level domain context discrimination. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 447–457. Association for Computational Linguistics.

Biao Zhang, Deyi Xiong, Jinsong Su, and Hong Duan. 2017. A context-aware recurrent encoder for neural machine translation. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, 25(12):2424–2432.

Jian Zhang, Liangyou Li, Andy Way, and Qun Liu. 2016. Topic-informed neural machine translation. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1807–1817. The COLING 2016 Organizing Committee.

Shiqi Zhang and Deyi Xiong. 2018. Sentence weighting for neural machine translation domain adaptation. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3181–3190. Association for Computational Linguistics.