Multilingual Neural Machine Translation

Raj Dabre[1]
NICT
Kyoto, Japan.

Chenhui Chu[2]
Kyoto University
Kyoto, Japan.

Anoop Kunchukuttan[3]
Microsoft
Hyderabad, India.

[1]raj.dabre@nict.go.jp, [2]chu@i.kyoto-u.ac.jp,
[3]anoop.kunchukuttan@microsoft.com

Abstract

The advent of neural machine translation (NMT) has opened up exciting research in building multilingual translation systems i.e. translation models that can handle more than one language pair. Many advances have been made which have enabled (1) improving translation for low-resource languages via transfer learning from high resource languages; and (2) building compact translation models spanning multiple languages. In this tutorial, we will cover the latest advances in NMT approaches that leverage multilingualism, especially to enhance low-resource translation. In particular, we will focus on the following topics: modeling parameter sharing for multi-way models, massively multilingual models, training protocols, language divergence, transfer learning, zero-shot/zero-resource learning, pivoting, multilingual pre-training and multi-source translation.

1 Relevance to CL community

Machine translation (MT) is one of the most challenging problems in CL and AI, and MT research has been at the forefront of many advances in the field. Since its advent in 2014, neural machine translation (NMT) (Sutskever et al., 2014; Bahdanau et al., 2015) has become the dominant paradigm and has shown the benefits of deep learning for NLP. While initial research on NMT started with building translation systems between two languages, researchers discovered that the NMT framework can naturally incorporate multiple languages. We refer to NMT systems handling translation between more than one language pair as *multilinugal NMT* (MNMT) systems.

There are multiple use cases and benefits for MNMT systems: (1) improving translation for low-resource languages via transfer learning from high-resource languages; (2) better generalization from exposure to diverse languages; (3) building compact translation models spanning multiple languages; (4) rapidly building MT systems by adapting existing multilingual models. In the past few years, there has been a lot of research addressing these themes and the area continues to be actively researched. Hence, it would be timely to have a tutorial which systematically presents the cutting-edge work in the area of MNMT. This would be of interest to researchers and practitioners of MT.

More broadly, multilingual NLP has received a lot of interest in recent times driven by two important questions:

Q1. *How do we build distributed representations such that similar text across languages have similar representations?*

Q2. *Is it possible to have a one-model-for-all-languages solution to NLP applications despite lacking data for certain languages?*

We believe that MNMT is a natural starting point to investigate these two important questions for NLP research. Hence, the tutorial would also be of interest to researchers and practitioners who work on multilingual NLP.

Proceedings of the 28th International Conference on Computational Linguistics: Tutorial Abstracts, pages 16–21
Barcelona, Spain (Online), December 12th, 2020.

2 Tutorial Overview

The tutorial will draw material from a survey paper on multilingual NMT that we have authored (Dabre et al., 2020). This paper has been published in ACM Computing Surveys. We also intend to cover some latest advances that are not mentioned in the survey paper. We will divide the tutorial in two parts, the first focusing on general purpose multilingual modeling and the second focusing on multiple usecases for multilingual NMT.

In the first part, we will first present an overview of the basics of NMT and cross-lingual embeddings. We will establish the fundamentals of MNMT and focus on various design choices. Design choices will involve network architecture, training protocols, data processing, hyper-parameter tuning so that they can successfully incorporate multilingualism. We will discuss specific changes to be made for translation between related languages as well as unrelated languages. We will also talk about the limits of massively MNMT models and provide a cost-benefit analysis from the perspective of deploying such models.

In the latter half of the tutorial, we will first focus on the challenging scenario of translation between language pairs for which there are few or no parallel corpora. We will introduce various ways to leverage pivot languages and on transfer learning approaches. We will show how transfer learning approaches such as fine-tuning and teacher-student learning can be optimally done when language relatedness and syntax are explicitly addressed. We will also touch upon unsupervised NMT which addresses low-resource MT using just monolingual corpora and is complimentary to NMT. We will see how multilingualism and unsupervised approaches can be combined. We will see how MT models for new languages can be rapidly adapted from pre-trained MNMT models. Additionally, we will spend some time on multi-source NMT which leverages multilingual redundancy in terms of input in order to yield high quality translations. We will end the tutorial with a discussion on possible future directions that we believe that MNMT research should take.

3 Tutorial Outline

Some representative papers are mentioned against tutorial sections.

1. Introduction (15 min)
   - Why MNMT?
   - Motivating multilingual NLP
   - Cross-lingual embeddings (Conneau et al., 2018; Jawanpuria et al., 2019)
   - Tutorial roadmap
2. Basics of NMT (20 min) (Sutskever et al., 2014; Bahdanau et al., 2015; Sennrich et al., 2016b; Sennrich et al., 2016a; Vaswani et al., 2017)
   - Architectures (RNN/Transformer)
   - Pre-processing and training
   - Decoding (beam-search, reranking)
3. Multi-way translation (45min) (Firat et al., 2016a; Johnson et al., 2017)
   - Prototypical architectures
   - Controlling parameter sharing (Sachan and Neubig, 2018; Platanios et al., 2018; Wang et al., 2018)
   - Addressing language divergence (Vázquez et al., 2018; Gu et al., 2018)
   - Training protocols (Tan et al., 2019; Lakew et al., 2018)
   - Massively multilingual models (Aharoni et al., 2019; Bapna et al., 2019)
4. – Coffee Break –

5. Transfer learning (30min)
   - Fine-tuning approaches (Zoph et al., 2016; Firat et al., 2016b)
   - Utilizing language relatedness (Dabre et al., 2017b; Kocmi and Bojar, 2018)
   - Lexical and syntactic transfer (Nguyen and Chiang, 2017; Murthy et al., 2019)
   - Rapid adaption of MT models (Neubig and Hu, 2018; Gheini and May, 2019)
6. Zero-shot/zero-resource learning (30 min) (Johnson et al., 2017; Firat et al., 2016b; Cheng et al., 2017)
   - Pivoting strategies
   - Modified training objectives (Al-Shedivat and Parikh, 2019)
   - Teacher-student learning (Chen et al., 2017)
   - Unsupervised learning (Lample et al., 2018; Xia et al., 2019; Sen et al., 2019)
7. Multi-source translation (15 min) (Zoph and Knight, 2016; Dabre et al., 2017a)
   - Missing sentences (Nishimura et al., 2018)
   - Hybrid multi-source systems
   - Post-editing (Chatterjee et al., 2017)
8. Future directions (15 min)
9. Summary and conclusion (10 min)

Total Time: 180 minutes (excluding break).
Type of the Tutorial: Cutting-edge.
Pre-requisites: Familiarity with sequence to sequence learning.

4 Tutorial Instructors

Dr. Raj Dabre

Post-Doctoral Researcher, NICT, Kyoto, Japan
email: raj.dabre@nict.go.jp
website: https://prajdabre.wixsite.com/prajdabre

Raj Dabre received his M.Tech. from IIT Bombay, India and his Ph.D. from Kyoto University, Japan. He is a post-doctoral researcher at NICT which is Japan's national research institution for communication technologies. His research interests center on natural language processing, particularly neural machine translation for low resource languages and on model compression and computing efficiency. He has MT-related publications in ACL, EMNLP, AAAI, NAACL, COLING, INTERSPEECH and WMT. He was a member of the organizing committee of COLING 2012, is a current member of the organizing committee of the Workshop on Asian Translation and has coordinated joint research between Kyoto University (Japan) and IIT Bombay (India).

Dr. Chenhui Chu

Program-Specific Associate Professor, Graduate School Of Informatics, Kyoto University, Japan.
email: chu@i.kyoto-u.ac.jp
website: http://researchmap.jp/chu/?lang=english

Chenhui Chu received his B.S. in Software Engineering from Chongqing University in 2008, and M.S., and Ph.D. in Informatics from Kyoto University in 2012 and 2015, respectively. He is currently a program-specific associate professor at Kyoto University. His research won the MSRA collaborative research 2019 grant award, 2018 AAMT Nagao award, and CICLing 2014 best student paper award. He is on the editorial board of the Journal of Natural Language Processing, Journal of Information

Processing, and a steering committee member of Young Researcher Association for NLP Studies. His research interests center on natural language processing, particularly machine translation and language and vision understanding.

Dr. Anoop Kunchukuttan

Senior Applied Researcher, Microsoft, Hyderabad, India.
email: anoop.kunchukuttan@microsoft.com
website: http://anoopk.in
Anoop Kunchukuttan is a Senior Applied Researcher in the MT team at Microsoft India, where he is involved in building MT systems for multiple Indian languages. His research interests span in different aspects of machine translation, particularly: multilingual models, low-resource translation and translation involving related languages. More broadly, he is interested in different multilingual, cross-lingual and multi-task NLP approaches. He is passionate about building software and resources for NLP in Indian languages. He is the developer of the *Indic NLP Library* (Kunchukuttan, 2020), co-developer of the IndicNLP-Suite (Kakwani et al., 2020) and a co-founder of the AI4Bharat-NLP Initiative, a community initiative to build Indian language NLP technologies. He has published papers on MT and multilingual learning at ACL, NAACL, EMNLP, TACL and IJCNLP. He has been a member of the organizing committees for COLING 2012 and Workshop on Asian Translation. He received his Ph.D from the Indian Institute of Technology Bombay.

References

Roee Aharoni, Melvin Johnson, and Orhan Firat. 2019. Massively Multilingual Neural Machine Translation. In *NAACL*.

Maruan Al-Shedivat and Ankur Parikh. 2019. Consistency by agreement in zero-shot neural machine translation. In *NAACL*.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *In Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*, San Diego, USA, May. International Conference on Learning Representations.

Ankur Bapna, Colin Andrew Cherry, Dmitry (Dima) Lepikhin, George Foster, Maxim Krikun, Melvin Johnson, Mia Chen, Naveen Ari, Orhan Firat, Wolfgang Macherey, Yonghui Wu, Yuan Cao, and Zhifeng Chen. 2019. Massively multilingual neural machine translation in the wild: Findings and challenges. *arXiv:1907.05019*.

Rajen Chatterjee, M. Amin Farajian, Matteo Negri, Marco Turchi, Ankit Srivastava, and Santanu Pal. 2017. Multi-source neural automatic post-editing: Fbk's participation in the WMT 2017 ape shared task. In *Proceedings of the Second Conference on Machine Translation*, pages 630–638. Association for Computational Linguistics.

Yun Chen, Yang Liu, Yong Cheng, and Victor O.K. Li. 2017. A teacher-student framework for zero-resource neural machine translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1925–1935. Association for Computational Linguistics.

Yong Cheng, Qian Yang, Yang Liu, Maosong Sun, and Wei Xu. 2017. Joint training for pivot-based neural machine translation. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 3974–3980.

Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2018. Word translation without parallel data. In *Proceedings of the International Conference on Learning Representations*. URL: `https://github.com/facebookresearch/MUSE`.

Raj Dabre, Fabien Cromieres, and Sadao Kurohashi. 2017a. Enabling multi-source neural machine translation by concatenating source sentences in multiple languages. In *Proceedings of MT Summit XVI, vol.1: Research Track*, pages 96–106.

Raj Dabre, Tetsuji Nakagawa, and Hideto Kazawa. 2017b. An empirical study of language relatedness for transfer learning in neural machine translation. In *Proceedings of the 31st Pacific Asia Conference on Language, Information and Computation*, pages 282–286. The National University (Phillippines).

Raj Dabre, Chenhui Chu, and Anoop Kunchukuttan. 2020. A Survey of Multilingual Neural Machine Translation. *ACM Computing Surveys*, 53(5), September.

Orhan Firat, Kyunghyun Cho, and Yoshua Bengio. 2016a. Multi-way, multilingual neural machine translation with a shared attention mechanism. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 866–875. Association for Computational Linguistics.

Orhan Firat, Baskaran Sankaran, Yaser Al-Onaizan, Fatos T. Yarman Vural, and Kyunghyun Cho. 2016b. Zero-resource translation with multi-lingual neural machine translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 268–277. Association for Computational Linguistics.

Mozhdeh Gheini and Jonathan May. 2019. A universal parent model for low-resource neural machine translation transfer. *arXiv:1909.06516*.

Jiatao Gu, Hany Hassan, Jacob Devlin, and Victor O.K. Li. 2018. Universal neural machine translation for extremely low resource languages. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 344–354. Association for Computational Linguistics.

Pratik Jawanpuria, Arjun Balgovind, Anoop Kunchukuttan, and Bamdev Mishra. 2019. Learning multilingual word embeddings in latent metric space: a geometric approach. *Transaction of the Association for Computational Linguistics (TACL)*.

Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, Gokul N.C., Avik Bhattacharyya, Mitesh M. Khapra, and Pratyush Kumar. 2020. IndicNLPSuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for Indian languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4948–4961, Online, November. Association for Computational Linguistics.

Tom Kocmi and OndÅ™ej Bojar. 2018. Trivial transfer learning for low-resource neural machine translation. In *Proceedings of the Third Conference on Machine Translation, Volume 1: Research Papers*, pages 244–252, Belgium, Brussels, October. Association for Computational Linguistics.

Anoop Kunchukuttan. 2020. The IndicNLP Library. `https://github.com/anoopkunchukuttan/indic_nlp_library/blob/master/docs/indicnlp.pdf`.

Surafel Melaku Lakew, Aliia Erofeeva, Matteo Negri, Marcello Federico, and Marco Turchi. 2018. Transfer learning in multilingual neural machine translation with dynamic vocabulary. In *IWSLT*.

Guillaume Lample, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. 2018. Unsupervised machine translation using monolingual corpora only. In *International Conference on Learning Representations*.

V. Rudra Murthy, Anoop Kunchukuttan, and Pushpak Bhattacharyya. 2019. Addressing word-order divergence in multilingual neural machine translation for extremely low resource languages. In *NAACL*.

Graham Neubig and Junjie Hu. 2018. Rapid adaptation of neural machine translation to new languages. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 875–880. Association for Computational Linguistics.

Toan Q. Nguyen and David Chiang. 2017. Transfer learning across low-resource, related languages for neural machine translation. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 296–301. Asian Federation of Natural Language Processing.

Yuta Nishimura, Katsuhito Sudoh, Graham Neubig, and Satoshi Nakamura. 2018. Multi-source neural machine translation with data augmentation. In *15th International Workshop on Spoken Language Translation (IWSLT)*, Brussels, Belgium, October.

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.

Devendra Sachan and Graham Neubig. 2018. Parameter sharing methods for multilingual self-attentional translation models. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 261–271. Association for Computational Linguistics.

Sukanta Sen, Kamal Kumar Gupta, Asif Ekbal, and Pushpak Bhattacharyya. 2019. Multilingual unsupervised nmt using shared encoder and language-specific decoders. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016a. 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, Berlin, Germany, August. Association for Computational Linguistics.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016b. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany, August. 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*, NIPS'14, pages 3104–3112, Cambridge, MA, USA. MIT Press.

Xu Tan, Yi Ren, Di He, Tao Qin, and Tie-Yan Liu. 2019. Multilingual neural machine translation with knowledge distillation. In *International Conference on Learning Representations*.

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.

Raúl Vázquez, Alessandro Raganato, Jörg Tiedemann, and Mathias Creutz. 2018. Multilingual NMT with a language-independent attention bridge. *CoRR*, abs/1811.00498.

Yining Wang, Jiajun Zhang, Feifei Zhai, Jingfang Xu, and Chengqing Zong. 2018. Three strategies to improve one-to-many multilingual translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2955–2960. Association for Computational Linguistics.

Mengzhou Xia, Xiang Kong, Antonios Anastasopoulos, and Graham Neubig. 2019. Generalized data augmentation for low-resource translation. In *ACL*.

Barret Zoph and Kevin Knight. 2016. Multi-source neural translation. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 30–34. Association for Computational Linguistics.

Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. 2016. Transfer learning for low-resource neural machine translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 1568–1575.