WikipediA

Neural machine translation

Neural machine translation (NMT) is an approach to machine translation that uses a large artificial neural network to predict the likelihood of a sequence of words, typically modeling entire sentences in a single integrated model.

Deep neural machine translation is an extension of neural machine translation. Both use a large neural network with the difference that deep neural machine translation processes multiple neural network layers instead of just one.[1]

Contents

Properties

They require only a fraction of the memory needed by traditional statistical machine translation (SMT) models. Furthermore, unlike conventional translation systems, all parts of the neural translation model are trained jointly (end-to-end) to maximize the translation performance.[2][3][4]

History

Deep learning applications appeared first in speech recognition in the 1990s. The first scientific paper on using neural networks in machine translation appeared in 2014, followed by a lot of advances in the following few years. (Large-vocabulary NMT, application to Image captioning, Subword-NMT, Multilingual NMT, Multi-Source NMT, Character-dec NMT, Zero-Resource NMT, Google, Fully Character-NMT, Zero-Shot NMT in 2017) In 2015 there was the first appearance of a NMT system in a public machine translation competition (OpenMT'15). WMT'15 also for the first time had a NMT contender; the following year it already had 90% of NMT systems among its winners.[5]

Workings

NMT departs from phrase-based statistical approaches that use separately engineered subcomponents.[6] Neural machine translation (NMT) is not a drastic step beyond what has been traditionally done in statistical machine translation (SMT). Its main departure is the use of vector representations ("embeddings", "continuous space representations") for words and internal states. The structure of the models is simpler than phrase-based models. There is no separate language model, translation model, and reordering model, but just a single sequence model that predicts one word at a time. However, this sequence prediction is conditioned on the entire source sentence and the entire already produced target sequence.[7]

NMT models use deep learning and representation learning.

The word sequence modeling was at first typically done using a recurrent neural network (RNN). A bidirectional recurrent neural network, known as an *encoder*, is used by the neural network to encode a source sentence for a second RNN, known as a *decoder*, that is used to predict words in the target language.[8]

Convolutional Neural Networks (Convnets) are in principle somewhat better for long continuous sequences, but were initially not used due to several weaknesses that were successfully compensated for by 2017 by using so-called "attention"-based approaches.[9][10]

Usage

By 2016, most of the best MT systems were using neural networks.[5] Google, Microsoft and Yandex[11] translation services now use NMT. Google uses Google Neural Machine Translation (GNMT) in preference to its previous statistical methods.[12] Microsoft uses a similar technology for its speech translations (including Microsoft Translator live and Skype Translator).[13] An open source neural machine translation system, OpenNMT, has been released by the Harvard NLP group.[14] Yandex.Translator has a hybrid model: its translation provides a statistical model and a neural network. After this, the algorithm CatBoost, which is based on machine learning, will select the best of the obtained results[11] Machine translation providers who also offer neural machine translation include Pangeanic's language lab PangeaMT [15] , Omniscien Technologies (formerly Asia Online),[16] Tilde,[17] KantanMT,[18] SDL,[19] Globalese,[20] and TransPerfect. DeepL offers a generic machine translation system with deep learning AI systems while Omniscien Technologies provides customized deep neural machine translation (Deep NMT) and Systran offers Pure Neural Machine Translation with deep neural networks.

References

1. "Deep Neural Machine Translation" (https://omniscien.com/deep-neural-machine-translation/). Omniscien Technologies. Retrieved 2017-11-08.

2. Kalchbrenner, Nal; Blunsom, Philip (2013). "Recurrent Continuous Translation Models" (http://www.aclweb.org/anthology/D13-1176). *Proceedings of the Association for Computational Linguistics*.

3. Sutskever, Ilya; Vinyals, Oriol; Le, Quoc Viet (2014). "Sequence to sequence learning with neural networks". arXiv:1409.3215 (https://arxiv.org/abs/1409.3215) [cs.CL (https://arxiv.org/archive/cs.CL)].

4. Kyunghyun Cho; Bart van Merrienboer; Dzmitry Bahdanau; Yoshua Bengio (3 September 2014). "On the Properties of Neural Machine Translation: Encoder–Decoder Approaches". arXiv:1409.1259 (https://arxiv.org/abs/1409.1259) [cs.CL (https://arxiv.org/archive/cs.CL)].

5. Bojar, Ondrej; Chatterjee, Rajen; Federmann, Christian; Graham, Yvette; Haddow, Barry; Huck, Matthias; Yepes, Antonio Jimeno; Koehn, Philipp; Logacheva, Varvara; Monz, Christof; Negri, Matteo; Névéol, Aurélie; Neves, Mariana; Popel, Martin; Post, Matt; Rubino, Raphael; Scarton, Carolina; Specia, Lucia; Turchi, Marco; Verspoor, Karin; Zampieri, Marcos (2016). "Findings of the 2016 Conference on Machine Translation" (https://cris.fbk.eu/retrieve/handle/11582/307240/14326/W16-2301.pdf) (PDF). *ACL 2016 First Conference on Machine Translation (WMT16)*. The Association for Computational Linguistics: 131–198.

6. Wołk, Krzysztof; Marasek, Krzysztof (2015). "Neural-based Machine Translation for Medical Text Domain. Based on European Medicines Agency Leaflet Texts". *Procedia Computer Science*. **64** (64): 2–9. doi:10.1016/j.procs.2015.08.456 (https://doi.org/10.1016%2Fj.procs.2015.08.456).

7. Philipp Koehn (2016-11-30). "The State of Neural Machine Translation (NMT)" (https://omniscien.com/state-neural-machine-translation-nmt/). Omniscien Technologies. Retrieved 2017-11-08.

8. Dzmitry Bahdanau; Cho Kyunghyun; Yoshua Bengio (2014). "Neural Machine Translation by Jointly Learning to Align and Translate". arXiv:1409.0473 (https://arxiv.org/abs/1409.0473) [cs.CL (https://arxiv.org/archive/cs.CL)].

9. Bahdanau, Dzmitry; Cho, Kyunghyun; Bengio, Yoshua (2014-09-01). "Neural Machine Translation by Jointly Learning to Align and Translate". arXiv:1409.0473 (https://arxiv.org/abs/1409.0473) [cs.CL (https://arxiv.org/archive/cs.CL)].

10. Coldewey, Devin (2017-08-29). "DeepL schools other online translators with clever machine learning" (https://techcru nch.com/2017/08/29/deepl-schools-other-online-translators-with-clever-machine-learning/). *TechCrunch*. Retrieved 2018-01-27.

11. "Yandex — Company blog — One model is better than two. Yandex.Translate launches a hybrid machine translation system" (https://yandex.com/company/blog/one-model-is-better-than-two-yu-yandex-translate-launches-a-hybrid-mac hine-translation-system/). *Yandex*. Retrieved 2018-01-27.

12. Lewis-Kraus, Gideon (December 14, 2016). "The Great A.I. Awakening" (https://www.nytimes.com/2016/12/14/magaz ine/the-great-ai-awakening.html). The New York Times. Retrieved 2016-12-21.

13. "Microsoft Translator launching Neural Network based translations for all its speech languages" (https://blogs.msdn.mi crosoft.com/translation/2016/11/15/microsoft-translator-launching-neural-network-based-translations-for-all-its-speech -languages/). *Translator*. Retrieved 2018-01-27.

14. "OpenNMT – Open-Source Neural Machine Translation" (http://opennmt.net/). *opennmt.net*. Retrieved 2017-03-22.

15. *Machine Translation* | pangeamt.com https://pangeamt.com | pangeamt.com (https://pangeamt.com) Check |url= value (help). Retrieved 2018-11-12. Missing or empty |title= (help)

16. "Omniscien Technologies Announces Release of Language Studio with Next-Generation Neural Machine Translation Technology" (https://omniscien.com/omniscien-technologies-announces-release-language-studio-next-generation-neu ral-machine-translation-technology/). Omniscien Technologies. 2017-04-21. Retrieved 2017-11-08.

17. "MT Features: Neural Machine Translation | tilde.com" (https://tilde.com/products-and-services/machine-translation/fe atures/neural-translation). *Machine Translation*. Retrieved 2018-02-09.

18. "KantanMT – Cloud-based Machine Translation Platform" (https://www.kantanmt.com/displayarticle.php?id=WMA-kyc AAP8GvykS&slug=kantanmt-users-can-now-customise-and-deploy-neural-machine-translation-engines). *KantanMT*. Retrieved 2017-11-08.

19. "SDL Brings Neural Machine Translation to its Secure Enterprise Translation Server" (http://www.sdl.com/about/news-media/press/2017/sdl-brings-nmt-to-its-secure-ets.html). *SDL*. Retrieved 2017-11-08.

20. "Globalese 3.0 released" (http://www.globalese-mt.com/2017/09/05/globalese-3-0-released/). *Globalese*. 2017-09-05. Retrieved 2017-11-08.