Procesamiento del Lenguaje Natural, Revista nº 63, septiembre de 2019, pp. 151-154 recibido 31-03-2019 revisado 25-04-2019 aceptado 17-05-2019

Unsupervised Neural Machine Translation, a new paradigm solely based on monolingual text

Traducción Automática Neuronal no Supervisada, un nuevo paradigma basado solo en textos monolingües

Mikel Artetxe, Gorka Labaka, Eneko Agirre
IXA NLP Group, HiTZ Basque Center for Language Technologies
University of the Basque Country (UPV/EHU)
{mikel.artetxe, gorka.labaka, e.agirre}@ehu.eus

Abstract: This article presents UnsupNMT, a 3-year project of which the first year has already been completed. UnsupNMT proposes a radically different approach to machine translation: unsupervised translation, that is, translation based on monolingual data alone with no need for bilingual resources. This method is based on deep learning of temporal sequences and uses cutting-edge interlingual word representations in the form of cross-lingual word embeddings. This project is not only a highly innovative proposal but it also opens a new paradigm in machine translation which branches out to other disciplines, such us transfer learning. Despite the current limitations of unsupervised machine translation, the techniques developed are expected to have great repercussions in areas where machine translation achieves worse results, such as translation between languages which have little contact, e.g. German and Russian.
Keywords: Machine Translation, Deep Learning, Word Embedding

Resumen: Este artículo presenta UnsupNMT, un proyecto de 3 años del que ha trascurrido la primera anualidad. UnsupNMT plantea un método radicalmente diferente de hacer traducción automática: la traducción no supervisada, es decir, basada exclusivamente en textos monolingües sin ningún recurso bilingüe. El método propuesto se basa en aprendizaje profundo de secuencias temporales combinado con los últimos avances en representación interlingual de palabras ("cross-lingual word embeddings"). Además de ser una propuesta propiamente innovadora, abre un nuevo paradigma de traducción automática con ramificaciones en otras disciplinas como el aprendizaje por transferencia ("transfer learning"). A pesar de las limitaciones actuales de la traducción automática no-supervisada, se espera que las técnicas desarrolladas tengan gran repercusión en áreas donde la traducción automática consigue peores resultados, como la traducción entre pares de idiomas con poco contacto, tales como alemán o ruso.
Palabras clave: Traducción Automática, Aprendizaje Profundo, Word Embeddings

1 Introduction

Machine translation has been one of the most prominent applications of artificial intelligence since the very beginnings of the field. In addition to its intrinsic interest given the difficulty and completeness of the problem, machine translation has a huge practical interest in our increasingly global word, as it promises to break the language barrier while keeping the cultural heritage and diversity of all the languages spoken in the world.

In very recent times, previous approaches have been superseded by neural machine translation (NMT), which has now become the dominant paradigm to machine translation (Bahdanau, Cho, and Bengio, 2014). As opposed to the traditional statistical machine translation (SMT), NMT systems are trained end-to-end, take advantage of continuous representations that greatly alleviate the sparsity problem, and make use of much larger contexts, thus mitigating the locality problem. Thanks to this, NMT has been reported to significantly improve over SMT both in automatic metrics and human evaluation.

Current NMT methods require expensive annotated data, as they fail terribly when the training data is not big enough (Koehn

ISSN 1135-5948. DOI 10.26342/2019-63-18

and Knowles, 2017). Unfortunately, the lack of large parallel corpora is a practical problem for the vast majority of language pairs, including low-resource languages (e.g. Basque) as well as many combinations of major languages (e.g. German-Russian). Several authors have recently tried to address this problem using pivoting or triangulation techniques (Chen et al., 2017) as well as semi-supervised approaches (He et al., 2016), but these methods still require a strong cross-lingual signal.

In this project, we introduce unsupervised neural machine translation, a new paradigm where the system learns to translate between two languages without the need of any bilingual dictionary or translation memories. That is, given large bodies of monolingual text (monolingual corpora), the system is able to extract the patterns which allow to translate from one language to another.

Our approach would use standard deep learning models for sequence-to-sequence learning. More concretely, we would follow the encoder/decoder architecture, combining a single language independent encoder that would compose the cross-lingual word embeddings and several language specific decoders that would decompose this representation back into the appropriate language. The system would be trained in an unsupervised manner following the same principle of denoising autoencoders, and we would explore additional techniques like adversarial training and backtranslation to enhance proper learning (Sennrich, Haddow, and Birch, 2016).

The new techniques will open new research avenues on machine translation. We are at a very good position to check whether this new paradigm allows to improve the state-of-the-art in MT, especially for less-resourced language pairs and domains. Moreover, the developed methods to train cross-lingual sentence representations will also be useful for cross-lingual transfer learning, as already shown by cross-lingual word representations. Finally, the viability to induce translation models in a completely unsupervised environment would empirically prove the existence of an inherent connection among all languages, which has a great interest from the point of view of Linguistics.

2 Goals

The overall goal of this project is to develop unsupervised learning methods for neural machine translation. In order to maximize impact of this nascent technology, the project will be structured in three goals:

Goal 1: Develop methods to train NMT models in a completely unsupervised manner, relying solely in monolingual corpora. This is the core goal of this project. The rest of the goals explore practical ramifications and impact of unsupervised NMT.

Goal 2: Improve the state-of-the-art in machine translation in different real-world scenarios where we have access to varying degrees of cross-lingual signals. This goal explores whether the new paradigm has practical applications.

Goal 3: Transfer learning. This goal explores whether the new paradigm has practical implications in transferring natural language processing systems from a resource-rich language to a less-resourced language.

3 Technical approach

We propose a radically new approach to unsupervised machine translation based on deep learning, a direction that has shown to be highly successful in other related areas, including standard supervised machine translation itself through NMT. The core of our approach is to learn to compose cross-lingual word representations in an unsupervised manner. Then, use those embeddings to generate a first translation system that can be improved by monolingual techniques, such as denoising autoencoders or backtranslation.

Regarding the first step, and in order to obtain our cross-lingual word representations, we will rely on techniques pioneered by us to build cross-lingual embedding mappings (Artetxe, Labaka, and Agirre, 2017; Artetxe, Labaka, and Agirre, 2018a). Recent works in unsupervised word embedding mapping (Lample et al., 2018a) have managed to obtain results comparable to previous supervised techniques, which we managed to improve (Artetxe, Labaka, and Agirre, 2018b). However, existing methods are based on the geometric interpretation of the embedding space (e.g. minimizing the Euclidean distance, maximizing the cosine similarity...), which is unnatural and presumably suboptimal for machine translation. For that reason, we will explore alternative interpretations. At the same time, while existing models are bilingual, we plan to extend them to the multilingual scenario, so we can exploit the relationship among several languages at the same time.

These cross-lingual word-embeddings contain enough bilingual information to generate a rudimentary word-by-word translation system, which can be improved by different techniques. Either by directly generating a neural translation system that makes use of these pre-trained embeddings and is trained using techniques such as denoising autoencoders or backtranslation, or, taking advantage of the SMT modular architecture, by using these embeddings to generate a phrase-table that can be combined with a language model.

In relation to Goal 2, the project will explore extensions of the previous approach to exploit cross-lingual signals of different degree when available, which would be used to improve the state-of-the-art in machine translation in different practical scenarios. In particular, we plan to:

- If a small parallel corpus is available, use it to fine-tune our model taking care not to overfit. This type of fine-tuning has already been shown to be effective in NMT in the case of domain adaptation (Chu, Dabre, and Kurohashi, 2017).
- If a comparable corpus is available, use our model to iteratively extract reliable parallel sentences from it and improve the model with these parallel sentences.

While the main scenario under consideration in this project is that of machine translation, we also plan to explore the application of the developed methods in cross-lingual transfer learning (Goal 3), where a model is trained in one language and used in a different one. This has a great practical interest, as it allows to leverage the largely available annotated data and resources in major languages (in particular, English), to other less resourced ones (e.g. Basque or even Spanish for some tasks like co-reference, sentiment analysis or named-entity recognition). For that purpose, we plan to extend the existing methods based on cross-lingual word embeddings to incorporate the entire encoder learned with our approach, so the resulting system does not only account for word-level relations but also for more complex phrase- or sentence-level relations.

4 Current progress

The first attempts which obtained promising results in standard machine translation benchmarks using monolingual corpora only (Artetxe et al., 2018; Lample et al., 2018a) build upon unsupervised cross-lingual embedding mappings, which independently train word embeddings in two languages and learn a linear transformation to map them to a shared space. The resulting cross-lingual embeddings are used to initialize a shared encoder for both languages, and the entire system is trained using a combination of denoising autoencoding, back-translation and, in the case of (Lample et al., 2018a), adversarial training.

During the first months of the project we have made progress mainly in our core Goal 1, where we show that the modular architecture of phrase-based SMT is more suitable for this problem. Our work (Artetxe, Labaka, and Agirre, 2018c), concurrent with (Lample et al., 2018b), adapted the same principles discussed above to train an unsupervised SMT model, obtaining large improvements over the original unsupervised NMT systems. More concretely, we learn cross-lingual n-gram embeddings from monolingual corpora based on the mapping method discussed earlier, and use them to induce an initial phrase-table that is combined with an n-gram language model and a distortion model. This initial system is then refined through iterative back-translation.

More recently (Artetxe, Labaka, and Agirre, 2019), we identifiy and address several deficiencies of existing unsupervised SMT approaches by exploiting subword information, developing a theoretically well founded unsupervised tuning method, and incorporating a joint refinement procedure. Moreover, we use our improved SMT system to initialize a dual NMT model, which is further fine-tuned through on-the-fly back-translation. Together, we obtain large improvements over the previous state-of-the-art in unsupervised machine translation. For instance, we get 22.5 BLEU points in English-to-German WMT 2014, 5.5 points more than the previous best unsupervised system, and 0.5 points more than the (supervised) shared task winner back in 2014.

Acknowledgments

UnsupNMT is a project funded by the Spanish Ministry of Economy, Industry and Competitiveness (TIN2017-91692-EXP).

References

Artetxe, M., G. Labaka, and E. Agirre. 2017. Learning bilingual word embeddings with (almost) no bilingual data. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 451–462, Vancouver, Canada, July. Association for Computational Linguistics.

Artetxe, M., G. Labaka, and E. Agirre. 2018a. Generalizing and improving bilingual word embedding mappings with a multi-step framework of linear transformations. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, pages 5012–5019.

Artetxe, M., G. Labaka, and E. Agirre. 2018b. A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 789–798. Association for Computational Linguistics.

Artetxe, M., G. Labaka, and E. Agirre. 2018c. Unsupervised statistical machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3632–3642, Brussels, Belgium, October-November. Association for Computational Linguistics.

Artetxe, M., G. Labaka, and E. Agirre. 2019. An effective approach to unsupervised machine translation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics.

Artetxe, M., G. Labaka, E. Agirre, and K. Cho. 2018. Unsupervised neural machine translation. In *Proceedings of the 6th International Conference on Learning Representations (ICLR 2018)*, April.

Bahdanau, D., K. Cho, and Y. Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv e-prints*, abs/1409.0473, September.

Chen, Y., Y. Liu, Y. Cheng, and V. O. 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.

Chu, C., R. Dabre, and S. 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.

He, D., Y. Xia, T. Qin, L. Wang, N. Yu, T.-Y. Liu, and W.-Y. Ma. 2016. Dual learning for machine translation. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*. Curran Associates, Inc., pages 820–828.

Koehn, P. and R. Knowles. 2017. Six challenges for neural machine translation. In *Proceedings of the First Workshop on Neural Machine Translation*, pages 28–39. Association for Computational Linguistics.

Lample, G., A. Conneau, L. Denoyer, and M. Ranzato. 2018a. Unsupervised machine translation using monolingual corpora only. In *Proceedings of the 6th International Conference on Learning Representations (ICLR 2018)*, April.

Lample, G., M. Ott, A. Conneau, L. Denoyer, and M. Ranzato. 2018b. Phrase-based & neural unsupervised machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 5039–5049, Brussels, Belgium, October-November. Association for Computational Linguistics.

Sennrich, R., B. Haddow, and A. 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, Berlin, Germany, August. Association for Computational Linguistics.