ITAI: Adaptive Neural Machine Translation Platform

ITAI: Plataforma de traducción automática neuronal adaptativa

Thierry Etchegoyhen,[1] **David Ponce,**[1]
Harritxu Gete Ugarte,[1] **Victor Ruiz Gómez,**[1]
[1]Vicomtech Foundation, Basque Research and Technology Alliance (BRTA)
{tetchegoyhen,adponce,hgete,vruiz}@vicomtech.org

Abstract: We describe an adaptive neural machine translation platform which integrates continuous learning and supports multiple use-cases in the translation industry. The application is being developed and evaluated within the applied research project ITAI. Research within the project has shown the potential of the platform to cover the main identified use cases and provide rapid adaptation via continuous learning.
Keywords: Neural machine translation, Continuous learning

Resumen: Este artículo presenta una plataforma de traducción automática neuronal que integra aprendizaje continuo y da soporte a múltiples casos de uso en la industria de la traducción. La aplicación se está desarrollando y evaluando en el marco del proyecto de investigación aplicada ITAI. La investigación realizada resalta el potencial de la plataforma para dar soporte a los casos de uso identificados y proveer una rápida adaptación de las traducciones mediante aprendizaje continuo.
Palabras clave: Traducción automática neuronal, Aprendizaje continuo

1 Introduction

Neural Machine Translation (NMT) (Bahdanau, Cho, and Bengio, 2015; Vaswani et al., 2017) has brought significant gains in Machine Translation (MT) quality and has become the dominant paradigm in both academic research and commercial exploitation. This technology is being increasingly integrated in the translation industry to support growing translation needs in the digital era.

Providing adequate support to the translation industry requires taking two main aspects into account.

First, actual practices in the industry feature a wide array of scenarios depending on the IT infrastructure at hand and the network of translators working for a specific company in the field. Translation may thus be performed via computer-assisted translation (CAT) tools such as SDL Trados Studio or Wordfast, to name two of the main ones, in Content Management System (CMS) environments, or directly performed in document editors such as Libre Office or MS Word. This disparity makes it difficult to bring MT technology to a significant portion of the translation industry.

Secondly, MT systems are usually only updated periodically when significant volumes of new training data become available and, therefore, do not provide timely adaptation of MT output corrections generated via post-editing. This limitation can result in a loss of productivity and increased frustration on the part of translators tasked to repeatedly correct identical errors over time when querying MT engines. Continuous learning (CL) addresses this issue via continuous updates of MT models on the basis of post-edited machine translation output fed back to model training processes. In NMT, CL usually takes the form of Online Learning (OL), where each new pair of source sentence and post-edited translation is used to update the corresponding model (Peris and Casacuberta, 2019; Turchi et al., 2017; Wuebker, Simianer, and DeNero, 2018; Domingo et al., 2019), although CL could also be performed via micro-batches with slightly delayed integration of user feedback.

One key aspect in CL is determining the proper trade-off between rapid adaptation of the models from user corrections and model stability over time, a topic which has only

been partially explored so far. Optimal integration of CL methods remains a matter of active research and is of key importance to provide useful adaptive machine translation technology.

In this paper, we describe ITAI, an adaptive neural machine translation platform which integrates continuous learning and supports multiple use-cases in the translation industry. The platform is being developed within the applied research project ITAI, partially supported by the Department of Economic Development of the Basque Government (Spri). The project started in April 2019 and will finalise in December 2021. It is carried over by the following consortium: MondragonLingua[1] (project coordinator), Emun[2], iAmetza[3], Mixer[4], Tai Gabe[5] and Vicomtech[6]. The project takes into account the translation requirements from each company and continuous translators' feedback across development cycles.

2 ITAI

The architecture of ITAI is described in Figure 1. The application consists of the following main elements:

- Front-ends, from which users may interact with the application. The front-ends include a web-based user interface (UI), plugins for specific CAT tools, and supports CMS integration.
- A REST API, which exposes the functionality of the back-end and handles user authentication.
- A back-end, which includes the required components to perform machine translation, manage the data generated from the use of the system, and manage the training and selection of continuously updated NMT models.

The core workflow involves users requesting machine translation of texts or documents, post-editing the automated translations as needed, and sending validated translations to the system. These validated translations are fed back to the NMT models via continuous learning to produce incremental improvements of the models.

The ITAI UI is implemented in Angular and the back-end components in Go. We describe the main components and functionality in more detail in the following sections.

2.1 Front-ends

As noted in the introduction, translation activities in the industry cover a wide array of usage. To provide support for the main identified use cases, the application supports different entry points.

We first addressed the most commonly used frameworks for multilingual content generation. CAT tools are an important part of the translation ecosystem, and we developed a specific plugin to connect the popular SDL Trados environment to the application, similarly to Domingo et al. (2019). Other frameworks such as Wordfast Classic have also been configured and tested within the project. Other CAT environments with support for custom MT can be easily configured to interact with ITAI via its REST API. Additionally, ITAI supports integration within CMS environments, and specific developments are being carried out within the project for the Ubiquo[7] environment.

We also developed a Web-based user interface to provide an additional access point to the functionalities of the system. Such an environment was identified as necessary for two main reasons. First, translation is also being carried out professionally outside dedicated environments such as CAT tools and without any technological support to improve productivity. Secondly, some proprietary CAT tools do not support the transfer of post-edited translations to external applications and users' feedback cannot be reflected in the MT models. In either case, users are thus limited in their interaction with supporting MT technology.

To address these issues, the ITAI UI offers a full-fledged access to the MT technology supported by the back-end. Users with little or no access to MT technology may thus upload documents and retrieve automatically translated documents maintaining the original format. The translated documents can then be post-edited in an external environment and the resulting validated translations uploaded via the UI, where the content will

[1] https://www.mondragonlingua.com/en/
[2] https://www.emun.eus/en/
[3] https://iametza.eus/
[4] http://www.mixer.com.es/es/mixer/
[5] https://www.naiz.eus/
[6] https://www.vicomtech.org
[7] https://www.ubiquo.me/

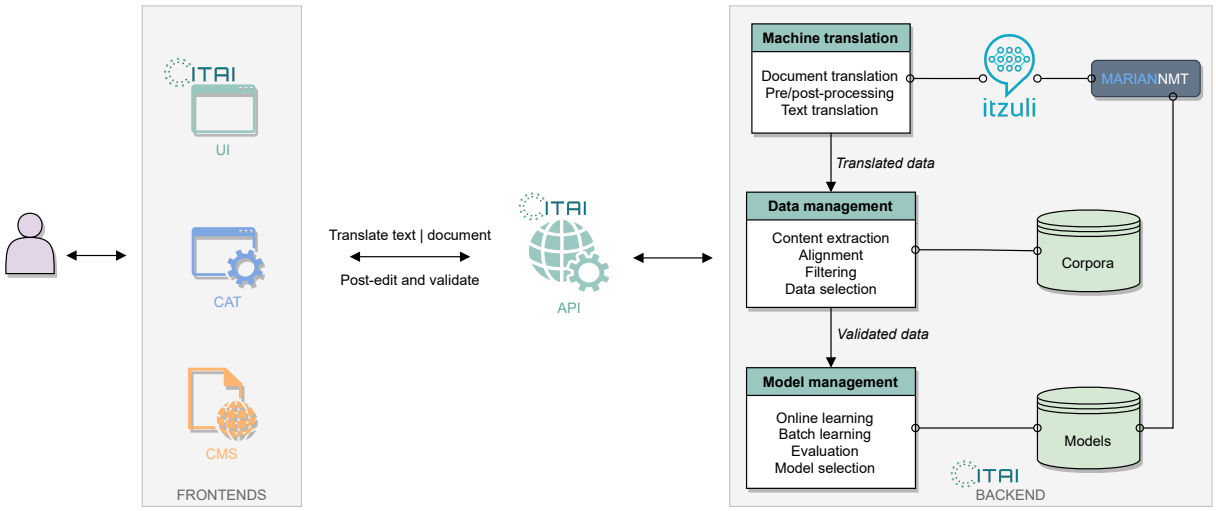

Figure 1: ITAI Architecture

be extracted to feed the MT models. Dual use-cases are also supported, where users may work with a CAT tool that supports the integration of ITAI MT services, and upload translation memories or documents containing their post-edited data via the ITAI UI. Additionally, the UI provides a simple editor where users can directly post-edit machine translated output that has been automatically segmented and filtered, as a default environment to correct and validate translations prior to sending them for integration in the MT models.

The user interface also provides dashboards to monitor volumes of translated and validated data, and list content that is pending validation to dissociate use of MT from the provision of post-edited data, as priorities usually differ for these two activities.

2.2 Back-end

The ITAI back-end provides support for three main types of functionality, which we describe in turn in the next sections.

2.2.1 Machine translation

Machine translation is carried out with Vicomtech's Itzuli MT toolkit, via its two main components: itzuli-translator and itzuli-doctrans.

The former performs text translation and can be deployed in scalable Kubernetes mode or as a standalone platform in a dedicated server. It integrates MarianNMT (Junczys-Dowmunt et al., 2018) to perform efficient NMT inference and training.

Document translation is done via itzuli-doctrans, a robust component for translation of documents in a variety of formats (odt, docx, xlsx, pptx, html or xliff, among others). It performs content extraction, text translation via itzuli-translator, and document reconstruction with format preservation.

Itzuli is a validated platform which supports large-scale translation services and provides ITAI with robust MT functionality.[8]

All NMT models currently deployed in ITAI are Transformer models (Vaswani et al., 2017), trained on large volumes of parallel, comparable and synthetic data. Although the platform is agnostic in terms of language pairs and domains, special emphasis is placed within the project on translation between Basque and English, French or Spanish, to contribute to improving language technology for the Basque language.

2.2.2 Data management

As one of the main goals of the project is to gradually increase the quality of MT models via continuous learning from user-generated corrections and validations, data management is a key functionality of the platform.

Since the platform allows for the provision of post-edited data via documents, in addition to the provision of segment-level data, the component supports sentence alignment with a combination of the metrics generated

[8]It notably supports the internal and public MT services of the Basque Government (https://www.euskadi.eus/traductor/), MondragonLingua's commercial MT services for Basque (https://lingua.eus/eu/itzultzailea) and domain-specific translation, and Vicomtech's public platform for the improvement of Basque translation technology (https://www.batua.eus/).

by the HunAlign (Varga et al., 2005) and STACC (Etchegoyhen and Azpeitia, 2016) aligners. It also performs several types of filtering, to identify misaligned or noisy data, via alignment scores and regular expression-based filters. Data selection is then performed to determine relevant data for continuous learning, given previous history.

Finally, the component also generates translation memories from validated aligned data, which users can download as a by-product of the data management processes.

2.2.3 Model management

Previously unseen data validated by users reach the model management component, where continuous learning takes place. New pairs, consisting of a source sentence and its validated translation, are used to adapt the relevant models, with a single update for online learning using the appropriate learning rate for the selected optimiser. Automatic evaluation then takes place to measure the impact of the update on both the new pairs and static test sets for the models at hand.

Although online learning is a relevant framework to adapt MT models on the fly, it is still an open matter to determine an optimal balance between aggressive adaptation, required for online learning to take effect on the basis of single data points, and model stability over time, necessary to maintain the overall quality of the models.

Several experiments are being carried out within the ITAI project to determine the appropriate configurations in this respect. Current results tend to favour a hybrid approach, with online learning performed for rapid MT adaptation useful to the users of the system, and batch fine-tuning over prior model training checkpoints once the volumes of accumulated new data reach a significant threshold. Model management processes for continuous learning will be adapted as necessary as final conclusions are reached within the project regarding continuous learning for neural machine translation.

3 Conclusions

In this paper, we described a neural machine translation platform which supports continuous learning and multiple use cases in the translation industry. The application is already operative within the applied research project ITAI and will be finalised in 2021. Research within the project has shown the potential of the platform to cover the main identified use cases and provide rapid adaptation via continuous learning. It also uncovered the need to further explore continuous learning for neural machine translation to reach an optimal balance between rapid adaptation and model stability over time.

References

Bahdanau, D., K. Cho, and Y. Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *Proc. of ICLR*.

Domingo, M., M. García-Martínez, A. Estela Pastor, L. Bié, A. Helle, Á. Peris, F. Casacuberta, and M. Herranz Pérez. 2019. Demonstration of a neural machine translation system with online learning for translators. In *Proc. of ACL*, pages 70–74.

Etchegoyhen, T. and A. Azpeitia. 2016. Set-Theoretic Alignment for Comparable Corpora. In *Proc. of ACL*, pages 2009–2018.

Junczys-Dowmunt, M., R. Grundkiewicz, T. Dwojak, H. Hoang, K. Heafield, T. Neckermann, F. Seide, U. Germann, A. Fikri Aji, N. Bogoychev, A. F. T. Martins, and A. Birch. 2018. Marian: Fast neural machine translation in C++. In *Proc. of ACL*, pages 116–121, July.

Peris, Á. and F. Casacuberta. 2019. Online learning for effort reduction in interactive neural machine translation. *Computer Speech & Language*, 58:98–126.

Turchi, M., M. Negri, A. Farajian, and M. Federico. 2017. Continuous learning from human post-edits for neural machine translation. *The Prague Bulletin of Mathematical Linguistics*, 108:233–244.

Varga, D., L. Németh, P. Halácsy, A. Kornai, V. Trón, and V. Nagy. 2005. Parallel corpora for medium density languages. In *Proc. RANLP*, pages 590–596.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 6000–6010.

Wuebker, J., P. Simianer, and J. DeNero. 2018. Compact personalized models for neural machine translation. In *Proc. of EMNLP*.