Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 3, No. 5, March : 2024

AI-ML AN EFFECTIVE TOOL FOR TRANSLATION OF TECHNICAL BOOKS IN REGIONAL LANGUAGES.

Er Binit Sarangi, Asst Director Factories and Boilors, Govt. of Odisha
Dr. Usharani Dash, Ex Professor Linguistic, Language specialist Odia and Santali and Translator, Odia, Odisha Bhasa Pratisthan, Govt of Odisha, Guest Faculty, Odia, AIIMS, Bhubaneswar
Dr Bimal Sarangi, Principal AIET, Bhubaneswar

Introduction : A qualified translator should always be used instead of machine translation. The recent initiative from AICTE, New Delhi and Govt. Of Odisha has been emphasizing to promote Odia language in transforming technical books from English to Odia through translation .This transformation definitely create a better way of communication of technical problems in very simple method ,so that there will be a better understanding of technology by the student and the stake holders.

However there is a bottleneckness for exact translation in short time. There by the help of AI and ML has to be incorporated as a quick translation method.

AI and ML as a translation tool :

In the present discourse on social change, machine translation (MT) is emerging as the poster child for a dystopian future in which artificial intelligence (AI) will replace highly skilled linguists and professional translators. Machine learning (MT) is one of the few fields in AI that has advanced to the point where using AI is becoming commonplace.

However, MT is only occasionally utilized for content translation in the professional language market, which is the realm of professional translators. Instead, the technology now helps people grasp materials written in a foreign language they are unfamiliar with on a basic level when combined with broad or generic language models. It addresses circumstances where one might have consulted.

Generic MT is not suitable for professional use

Commonly used and mostly free online resources like DeepL and Google Translate employ so-called generic machine translation (MT) models that are trained on the largest feasible data set. When it comes to start the process of converting manuscripts into common language—these tools can be of tremendous assistance.

Customized models enable vast increases in productivity :

The text's most crucial requirements are that the language be employed consistently; otherwise, a different picture would emerge. For instance, a Swiss Confederation research found that general machine translation (MT) "clearly reaches its limits when translating technical texts" and is therefore unsuitable for translating documents for publishing.

1. The study also found that while Deep-learning translations resulted in writings that were fluid in the target language,

The models must be tailored to particular requirements and circumstances in order to maximise the quality of the translation for machine translation (MT) to be effective in a professional setting. The selection and processing of the training data is what distinguishes these customized MT models from one another, rather than the technology employed.

Limiting the training data selection to a particular domain, such as solely utilizing linguistic data from the mechanical engineering field, enhances the models' ability to acquire technical words and apply them throughout the translation process. Even while it is already clear that the language quality has increased as a result, this method is unable to provide translations that are suitable for publication without additional editing.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 3, No. 5, March : 2024

Different methods of approach for Translation :

The experiment design compared the four different approaches detailed below. Approach D is the workflow applied at Lengoo, while approaches A, B, and C serve for comparison and correspond to other typical translation workflows:

A) Manual translation by a professional translator without using productivity tools

B) Manual translation by a professional translator using rule-based productivity tools Translation using a general machine translation model (such as GoogleTranslate or DeepL), followed by expert linguist post-editing

D) Translation performed by an MT model that has been specially trained using the client's translation memory,

thereafter having a qualified linguist post-edit the translation

Technical terminology must be learned by language models and used consistently :

Translation of extremely complicated materials, such technical documentation and construction manuals for pharmaceutical material, such as product labels or process descriptions, or on for maintenance instructions, demands a high level of uniformity and the accurate application of certain terminology. Since the integration of term bases and glossaries is more difficult in NMT technology than in rule-based systems, these demands on translation quality present serious obstacles.

The NMT models are not intended to translate previously translated information in an identical manner or to priorities any one word above others.

Currently, neural networks are used in machine learning (ML).

Every word is given a translation value that depends on the words that surround it. As a result, contextualization in the translation is made feasible by factorization, enabling NMT models to generate language that flows as naturally as possible and has a range of tones of voice. However, it is precisely this variety that prevents professional language use from being linguistically consistent, which is crucial.

Statistical machine translation received criticism for its grammatical monotony, strange linguistic flow, and frequent lack of logic throughout the first ten years of the twenty-first century. These days, NMT lacks precisely the benefits of this ostensibly outdated technology, which would enable its

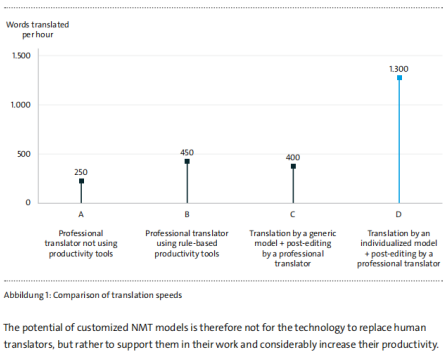

Abbildung 1: Comparison of translation speeds

The potential of customized NMT models is therefore not for the technology to replace human translators, but rather to support them in their work and considerably increase their productivity.

Fig -1

Translation quality can be improved through training with synthetic data.

Like all other fields where AI is used, translation is restricted in its application by two primary factors: the availability of high-quality data and the availability of computational capacity. It is not always assured that language data will be available, particularly when it comes to natural language. When dealing with language pairs that provide inadequate parallel data for neural network training, or what are referred to as low-resource language pairs, data scientists employ methods that can lessen the need on training data in both the target and source languages One of these approaches synthesizes training data artificially using a technique known as back translation. High-quality text in the target language is totally automatically translated backwards into the desired source language for this synthesis.

Method to be considered for application of AI and ML in NTM.

1. How much efficiency is gained in the process when employing AI as compared to the same process without AI?

In terms of client-specific MT, the experiment described in ↗**Section 3** clearly demonstrates how high this efficiency gain can be. Only if the increase in efficiency compared to existing solutions and software – i.e. the status quo – is high does it make sense to introduce AI into existing processes. In many other areas, however, rule-based software is still better suited.

2. How high is the acceptance of the technology among all those involved?

The introduction of new technology is rarely a guaranteed success, since human beings are and remain the decisive factor. The level of acceptance of new technology is a key driver for the potential success or failure of new technology. Therefore, users should already play a key role in the development phase, and the benefits for users should be communicated clearly Professional translators have been using technology to support their work for decades, helping them increase consistency of language and making it easy to find and use previously translated content. The introduction of NMT is fundamentally transforming the work of linguists; from translators who produce content themselves to proofreaders who check and stylistically improve previously translated content. This transformation must be managed, and professional translators trained in the use of the new technology.

3. How well can the solution be integrated into existing processes?

In addition to its acceptance by end users, the integration into already existing ecosystems plays a major role in the success of any technology. In the specific case of NMT, this means, for example, that companies that order translations from service providers are not required to adapt to new file formats or processes. Rather, the service provider needs to be able to integrate into the existing environment and the already proven processes on the customer's side. This allows the potential of the technology to be fully exploited and increases acceptance.

4. Are the companies and individuals involved »prepared« to set up feedback loops, to collect the necessary data on an ongoing basis and to thus guarantee the continuous training of the AI models?

Depending on the training data available, NMT can deliver great efficiency gains with just the initial training of the models. The great added value lies, however, in the continuous collection of high-quality data and regular training cycles. The feedback loop required for this, i.e. the constant feeding of data into the models, is key to the successful integration of AI into existing business processes. The resulting AI learning effect, which results in the continuous improvement of quality, means that the technology can deliver real added value.

The figure shows how many words per hour professional translators can translate when supported by continuously trained NMT models. As time progresses, it becomes very clear how efficiency increases each time the model is trained. This can collect additional metadata to improve the linguists' performance, such as the number of words translated per hour as shown in Figure (1). Innovations based on AI must be ready for use in the real world before they can be introduced.

However, since companies often have little or no experience with AI, it is advisable to introduce the technology in collaboration with a full-service partner with the necessary technological expertise, and which can accompany the client each step of the way.

Conclusion :

The global demand for translation has more than doubled over the last ten years. There are simply not enough professional translators in the world today to cope with the growing amount of content. It is thus obvious that the use of AI in the translation field could fill this supply gap, Particularly in light of how rapidly general language and automatic translation applications such as Google translate etc. closer look at the existing solutions shows that the use of AI in translation already offers great

advantages in the professional language market, but also places very complex demands on developers.

The term »Artificial Intelligence« often suggests an independent factor that is equal to human intelligence and on the same level. However, in the field of translation in particular it is clear that the use of machine learning – the most important component of AI – should be used above all to support humans. So, perhaps the terminology is, in fact, wrong, and we should rather be referring to »Augmented Intelligence«, i.e. the extension of human intelligence through technology. Although there are huge resources of available language data for the major world languages such as English and Spanish, as well as many other European and Asian languages, this is not the case for the vast majority of languages around the world. By using techniques such as fully automated back translation, it is possible to apply MT here, too. The potential of AI in this field, though, has a different starting point. UNESCO concludes that about 2,400 of the world's 7,000 independent languages are currently in danger, as only between a few hundred to some thousands of native speakers remain.7 For these languages, language models can be used to support the remaining native speakers, promote the active learning of the language, and hence preserve the language in the long term.

References :

1. Dinu et al. (2019). Training Neural Machine Translation To Apply Terminology Constraints. Retrieved April 10, 2020 from **↗https://www.aclweb.org/anthology/P19-1294.pdf**
2. Edunov, Ott et al (2018). Understanding Back-Translation at Scale. Retrieved April 10, 2020 from**https://arxiv.org/pdf/1808.09381v**
3. Kocmi and Bojar (2019). Transfer Learning across Languages from Someone Else's NMT Model. Retrieved April 10, 2020 from **↗https://arxiv.org/pdf/1909.10955.pdf**
4. Moseley, C. 2010. Atlas of the World's Languages in Danger, 3rd edn. Paris, UNESCO Publishing
5. Swiss Confederation (November 19, 2019). Report DeepL test. Retrieved April 5, 2020 from **↗https://uepo.de/wpontent/uploads/2019/12/schweiz_bericht_deeplest_2019.**
6.Statista**https://www.statista.com/statistics/257656/size-of-the-global-language-services-market/**