CMS - Intelligent Machine Translation with Adaptation and AI

Ruhul Amin[1], Mounika Mandapuram[2*]

[1]Senior Data Entry Control Operator (IT), ED-Maintenance Office, Bangladesh Bank (Head Office), Dhaka, **BANGLADESH**
[2]EKIN Solutions, 13800 Coppermine Rd, Herndon, VA 20171, **USA**

[*]Corresponding Contact:
Email: mounikamandapuram09@gmail.com

Manuscript Received: 11 Oct 2021 Accepted: 02 Dec 2021

ABSTRACT

Machine translation, an emerging breakthrough, has changed translation. Dictionary-based machine translation, computer-aided translation, and neural machine translation with AI as its fundamental technology have progressed. However, NMT with AI has advanced machine translation. Many translation concerns still need to be solved. Language changes with context and dialect. Artificial intelligence will enable systems with adaptive algorithms to collaborate with humans to translate content more efficiently and well. The human translation should be revised, according to some. All in all, human progress fixes faults. Neural networks in machine translation ensure that adaptive frameworks can interpret like human translators. AI still needs help with language training and translation. Given the diversity of linguistic patterns and civilizations, they address clever machines; even with AI, handling human productivity seems unlikely. Machine translation is an important NLP topic that uses computers and adaptive systems to understand standard dialects. Neural machine translation (NMT) has become the standard in real-world MT frameworks. We begin this study with a broad assessment of NMT strategies and discuss architecture, decoding, and data analysis to improve content translation. We then summarize the most helpful expert resources and tools. We conclude with a discussion of future paths.

Keywords: Adaptation, CMS, Machine Translation, Artificial Intelligence

INTRODUCTION

Internet translations sometimes use domain-specific phrases that require a domain-specific lexicon. Today, MT frameworks can read broad statements but not domain-specific terms. This depends on MT systems needing domain-specific data to learn from. The most popular translation technologies, such as Google Translate or open-source phrase-based MT systems trained on nonspecific data, are often used to understand content or particularly domain-specific materials, resulting in unclear translations. This issue is crucial for professional translators and browsers that store multilingual material and operate with documents from diverse domains and generic MT systems. However, manually applying these services takes time (Bodepudi et al., 2019). Discovering and installing domain-specific words in an MT system is vital for improving interpreter usefulness and translation quality in deeply specified areas. With advances in Artificial Intelligence (AI), which underpins Natural Language Processing (NLP), Machine Translation (MT) from one language to another has become more precise and natural. Machine Translation (MT) is best for managing expressions, not syllables, words, phrases, or sentences. The contextualizing issue persists. The possibilities of AI in translation and its limitations are exciting (Gutlapalli, 2016). In the current AI landscape, there are no plausible ideas on how to choose translation methods, connect translation tools, and meet everyone's needs and market positioning (Gutlapalli et al., 2019). This study then discusses a better way to interpret productively and viably with a mix of machine and human translation. It advances the structure of machine translation and human interpretation using AI methodologies and adaptive mechanisms.

Review of Literature

This section discusses language translation services and related work. These programs do not translate speech but use a similar encoder-decoder network as in this study. Since 2006, Google Translate has been the most helpful and smoothest online content translation service, with over 500 million users and 100 billion words translated daily. Initially, the organization employed Statistical Machine Translation (SMT) (Och et al., 1999) to evaluate words and expressions for translation trained on UN and European parliament documents and files. SMT employed predictive algorithms to translate texts, making sentence structure impossible. Finally, Google switched to a Neural Machine Translation (NMT) model architecture that could interpret sentences instead of words. Google Translate now supports 109 languages, up from a few initially. Speech-to-speech translation is also available using the 3-stage translation procedure (Luong et al., 2015).

While interpreting, the organization model finds patterns in many archives to find the most consistent word sequence in the target language with high precision (Gutlapalli, 2017c). The precision varies per language and has been questioned by experts, yet it is still the best translation model. Due to a valid concern for this service, we use Google Translate API, a publicly available library that can be incorporated and called for translation in raw Python code. Microsoft Translator is a cloud-based translation service that may be used by individuals or businesses (Guo et al., 2019). Its cloud-based speech translation API can provide language translation functionality for websites and mobile apps. Microsoft Translator API uses Neural Machine Translation by default. Microsoft Translator, often called Bing Translator, translated websites and texts like Google Translate. Skype translator provides mobile and desktop speech-to-speech independent translation using Microsoft Translator's Statistical Machine Translation technology. Skype Translator also translates instant messages into over 70 languages.

Machine Translation Evolves Throughout Time in Three Stages

Machine translation uses programming to convert text or speech from one language to another (Gutlapalli, 2016b). MT enhancement is linked to digital innovation, data analysis, and other technology. Dictionary-based machine translation, computer-aided translation, and neural machine translation with AI have all advanced over time.

Georgetown University and IBM completed the main English-Russian machine translation in 1954 to study the IBM-701 computer. This technique showed the usefulness of MT to society and logic, introducing machine translation research. Science and technology revolutionized the 1970s as logical and mechanical data interchange between countries increased, making language restrictions worse and requiring a solution. Manual translation was far from meeting these needs, and experts anticipated machines to do the arduous translation task.

After that, computer-aided translation (CAT) began in machine translation and swiftly expanded into digital machine translation. The CAT translation is different from the machine translation. Because it combines human translation with programmed translation, it does not depend on it. CAT programming uses translation memory, machine translation, and human translation. Since the 1980s, corpus-based translation has grown, and CAT has used machine translation methodologies to improve translation and phrasing information bases. Computers can help translators improve translation quality, skills with repetitious phrases, and the machine translation process (Mandapuram & Hosen, 2018). Memory database is essential to CAT translation. CAT can use the memory database to translate particular match sentences, but partial match sentences require human-assisted manual translation.

Neural Machine Translation is the latest computer translation innovation integrating AI frameworks to give more intelligent machine translation than the previous two. Natural Language Understanding, Natural Language Processing, Computer Machine Translation, Translation Memory, Statistical Machine Translation, and Deep Learning theories underpin its concept (Gutlapalli, 2017). Another theory is that it overcame machine translation innovation and achieved high quality and precision. It uses big data, data augmentation, and cloud computing as backend computing platforms, relies on profound learning developments, especially neural networks, gets different training information to the backend from online resources, and does feature mining with deep learning methods and quick training (Anastasiou & Gupta, 2011).

METHODOLOGY BEHIND THE PROCESSING OF MACHINE TRANSLATION (MT)

MT is trained by a pipeline that, in a nutshell, does the following: (I) recognizes, disambiguates, and connects all of the terms in the text; the terms and their connections are used to identify the domain of the text and channel out the terms that are not domain-specific; (iii) the translation of such terms is obtained by following the cross-lingual connections; and (iv) the bilingual domain-specific terms are inserted into the MT framework utilizing various techniques. In the following paragraphs, each step will have its detailed explanation.

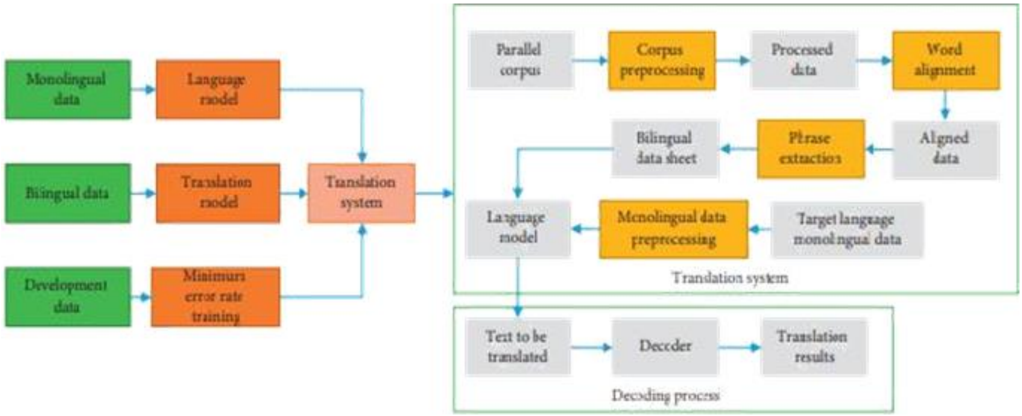

Figure 1: The Architecture of the Machine Translation System

This three-step pipeline uses statistical and machine learning techniques to train models using only open-source resources. Language training like stemming, morphological analysis, POS labeling, and parsing still needs to be done. This technique works with the framework's portability because it only needs a version with enough language and domain coverage. The first phase identifies and places terms by relevance using an essential statistical method based on tfidf weighting, where all n-grams from 1 to 10 are produced, and the IDF is found on web pages. The terms are linked to online text next. The linking step is a supervised word sense disambiguation issue that uses text resources to determine the proper sense and training information for each sentence, a list of expressions where the term appears. The application uses kernel-based word-master classifiers. The domain and syntagmatic components of term qualifying are shown using latent semantic and string kernels. The final phase improves connected phrases using data from open-source resources and pages. The pair's supplementary data includes elective terms (orthographic and morphological variations, synonyms, and related terms), photos, subjects, types, cross-language linkages, etc. The MT system must embed domain-specific bilingual terminology proof, per this analysis. The framework is expanded by filtering away concepts that do not fit a context.

To identify a sequence of phrases, we assign a domain to each linked term in a text, then find the most frequent domain and filter out out-of-scope terms. We cannot classify the entire content using typical text classification algorithms due to the large variety of languages and domains. An approach based on mapping categories to domains was devised. Human editors create and assign categories, making them less rigorous, clear, and consistent than ontologies. The cycle of the category hierarchy limits its usability. The domains are sequenced to make Natural Language Processing easier (Gutlapalli, 2017). Our strategy eliminates classification sparsity, reduces domains to 100,000, and requires no language-specific training data. This limits a large number of domains to one phrase. To determine the most relevant domain, these domains are ranked by occurrences. The most common domain is awarded to terms. Although manual mapping requires human participation, it is done once. It takes less time than explaining many training reports for text categorization because domain detection does not require reading the complete text.

Discovering the domain phrasing translation is the second last stage. We use the cross-language link to solve the page, not the word. To solve this, we proposed this method. We return the target page if the term matches the source page title or the most common form of the phrase in the target language.

A straightforward way to add bilingual terms to the MT framework is to link training information and terms. Despite outperforming more unexpected approaches, it has severe drawbacks that limit its applicability in real-world applications. When limited bilingual terms are linked to an extensive training dataset, terms with unclear translations are penalized because the most successive and general translations often get the highest probability, causing the MT framework to overlook specific translations (Mandapuram, 2016). This section discusses the Fill-Up and XML Markup techniques, which emphasize particular translations above generic ones. The Fill-Up model addresses a common circumstance when a large generic foundation model exists, and only a little in-domain information may be used to build an in-domain model. It will likely shield domain-specific information from in-domain information while using the foundation model's vast reach. Information connecting processes word categorization. Every corpus extracts and scores phrases separately. BYPASSING EXTERNAL INFORMATION, the XML markup method lets the decoder determine translations for specific source text ranges. Our source term identifies a range in the source text, while the target term is supplied to the decoder. Exclusive settings limit the decoder to specified translations and ignore other translation model translations.

SIGNIFICANT USES OF NEURAL MACHINE TRANSLATION IN VARIOUS VENTURES

Since MT and its growth are growing, more business sectors and models use its applications. This section examines real-world customer and B2B use cases that help improve machine translation applications.

Even if digital giants like Google Translate and Microsoft Translator deliver exact, consistent translations, some industries require precise training information to improve precision and proficiency. Because most MT models function on nonspecific information, conventional translators are no longer helpful or efficient. Small, medium, and large businesses use these use cases. Some companies offer bilingual translation services that are customizable for different domains. Different companies offer domain-specific translations. For effective translation, such technologies still require operator intervention.

Such machine translation applications translate text, voice, and image files from one language to another in real-time. Such generic translation procedures are more precise.

Facebook has extensively studied machine translation variables for nearly ten years. It published the Translations app (Mandapuram, 2017) to browse and translate its entire website using basic machine learning algorithms and publically approved text translations. Based on NMT, this app is clever and modern. Another comments-specific online translating service was released by Facebook in 2011. This tool's release announcement appears to be gone. However, Mashable (Mandapuram, 2017) reported this story in 2011.

Skype introduced speech translation in 2014 using cutting-edge speech recognition technology and MT to translate audio content between languages instantly. The organization says training information for machine translation and the model is extracted from transcribed pages, captioned video content, recently interpreted and translated one-on-one discussions, and organization information. The translation takes four easy stages. The framework converts sound to source language material using ASR (Automated Speech Recognition). It then standardizes content for speech translation. The framework uses machine learning-based book translation to translate the given language text into the suggested target language after standardization. Finally, the framework uses text-to-speech translation to make the sound.

Inaccuracies that may Result from the Use of Intelligent Machine Translation

The jury is out, especially since AI is tied to translation and languages. Despite warnings from significant participants in the field about the perils of unrestrained AI development, AI's immense benefits and advantages suggest that without a world-changing catastrophe, ongoing advancement is assured. Overall, we are not there. There has yet to be progress toward Strong AI (a conscious machine) or Artificial General Intelligence (a system that can apply its intelligence to multiple domains). Many domains need more progress toward human-level intelligence. In languages, the finest machine translation systems get outcomes about 80% of a qualified human translator.

People believe a neural network can do anything a person can with deep learning and enough training data, but this is false (Reddy et al., 2020). Despite its increasing accuracy, AI translation must catch up to human translators. Machine translation technologies and bilingual translators met to convert Korean to English and vice versa in a study. The 50-minute test (Mandapuram et al., 2018) found that 90% of NMT content needed to be more syntactically awkward or translated by a skilled local speaker. After their release, the highly publicized wearables "Google Buds" were praised for their machine translation accuracy, but assessments noted that they were too abnormal and inaccurate to be helpful. A real-time transitory translation may be helpful. However, an organization translating their website should have it translated by a human translator to connect with readers better than machines (Mandapuram et al., 2020). Online translation providers should address the simplicity of text extension. If your website looks great in English, you may need to create a distinct visual plan to accommodate longer words and phrases for accurate bilingual translation.

Conclusion

Along with the preceding enhancement, multi-task training can be arranged for future work. A list of source and destination language phonemes as expressions and written content is needed. The constraint significantly affected the results, according to the research. Additionally, movies dubbed and translated into multiple languages can be used for human voice model training. Many safety steps must be implemented to group source and target sentences and break them into single-sound expressions. A noise cancellation algorithm would also be needed to filter out background noise when actors spoke. AI machine translation is a good use of time since new ventures are finding ways to combine human and machine translations. Despite its name, it is not a catch-all app like Microsoft and Google. It is based mainly on AI but is meant to expand robots with people and vice versa, depending on the content and the suitable speed-precision trade-off. AI translation could succeed with human translators who can write and train them so robots cannot expose themselves. Only after noticing the innovation appears unsuitable for solo use.

REFERENCES

Anastasiou, D., & Gupta, R. (2011). Comparison of Crowdsourcing Translation with Machine Translation. Journal of Information Science, 37(6), 237-238.

Bodepudi, A., Reddy, M., Gutlapalli, S. S., & Mandapuram, M. (2019). Voice Recognition Systems in the Cloud Networks: Has It Reached Its Full Potential? *Asian Journal of Applied Science and Engineering*, *8*(1), 51–60. https://doi.org/10.18034/ajase.v8i1.12

Guo, J., Tan, X., He, D., Qin, T., Xu, L., and Liu, T. -Y. (2019). Non-autoregressive neural machine translation with enhanced decoder input. Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, p. 3723–3730

Gutlapalli, S. S. (2016). Commercial Applications of Blockchain and Distributed Ledger Technology. *Engineering International*, *4*(2), 89–94. https://doi.org/10.18034/ei.v4i2.653

Gutlapalli, S. S. (2017). An Early Cautionary Scan of the Security Risks of the Internet of Things. *Asian Journal of Applied Science and Engineering*, *6*, 163–168. Retrieved from https://ajase.net/article/view/14

Gutlapalli, S. S., Mandapuram, M., Reddy, M., & Bodepudi, A. (2019). Evaluation of Hospital Information Systems (HIS) in terms of their Suitability for Tasks. *Malaysian Journal of Medical and Biological Research*, *6*(2), 143–150. https://doi.org/10.18034/mjmbr.v6i2.661

Luong, T., Pham, H., and Manning, C. D. (2015). Effective approaches to attention-based neural machine translation, in Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, (Lisbon, Portugal), pp. 1412–1421, Association for Computational Linguistics.

Mandapuram, M. & Hosen, M. F. (2018). The Object-Oriented Database Management System versus the Relational Database Management System: A Comparison. *Global Disclosure of Economics and Business*, *7*(2), 89-96. https://doi.org/10.18034/gdeb.v7i2.657

Mandapuram, M. (2016). Applications of Blockchain and Distributed Ledger Technology (DLT) in Commercial Settings. *Asian Accounting and Auditing Advancement*, 7(1), 50–57. https://4ajournal.com/article/view/76

Mandapuram, M. (2017). Security Risk Analysis of the Internet of Things: An Early Cautionary Scan. *ABC Research Alert*, *5*(3), 49–55. https://doi.org/10.18034/ra.v5i3.650

Mandapuram, M., Gutlapalli, S. S., Bodepudi, A., & Reddy, M. (2018). Investigating the Prospects of Generative Artificial Intelligence *Asian Journal of Humanity, Art and Literature*, *5*(2), 167-174. https://doi.org/10.18034/ajhal.v5i2.659

Mandapuram, M., Gutlapalli, S. S., Reddy, M., Bodepudi, A. (2020). Application of Artificial Intelligence (AI) Technologies to Accelerate Market Segmentation. *Global Disclosure of Economics and Business*, *9*(2), 141-150. https://doi.org/10.18034/gdeb.v9i2.662

Och, F. J., Tillmann, C., and Ney, H. (1999). Improved alignment models for statistical machine translation, in 1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, 1999.

Reddy, M., Bodepudi, A., Mandapuram, M., & Gutlapalli, S. S. (2020). Face Detection and Recognition Techniques through the Cloud Network: An Exploratory Study. *ABC Journal of Advanced Research*, *9*(2), 103–114. https://doi.org/10.18034/abcjar.v9i2.660

--0--