DOI: 10.53469/jssh.2024.6(07).09

Research on Methods to Enhance Machine Translation Quality Between Low-Resource Languages and Chinese Based on ChatGPT

Jinyue Qi

School of Foreign Studies, Northwestern Polytechnical University, Xi'an, Shaanxi, China
2772885875@qq.com

Abstract: ***In recent years, machine translation engines have leveraged both traditional statistical models and the newer neural network models, achieving significant improvements in translation quality through the use of large-scale, high-quality corpora. These advancements have led to continuous improvements in translation quality for high-resource languages. However, the translation performance for low-resource languages remains suboptimal, primarily due to the difficulty in obtaining large-scale bilingual parallel corpora necessary for training neural network models. This study aims to enhance machine translation quality for low-resource languages by utilizing large language models, exploring various methods to improve translation quality, and evaluating their effectiveness. Specifically, the research focuses on comparing the effectiveness of these two methods through human evaluation using the Multidimensional Quality Metrics (MQM) framework.***

Keywords: Low-resource languages, Pivot Translation Technique, Pivot Prompting, Machine translation quality.

1. Introduction

In recent years, with the rapid development in the field of machine translation, many machine translation engines have utilized traditional statistical machine translation models and emerging neural machine translation models. Through large-scale, high-quality corpus training, there have been significant improvements and enhancements in translation quality. The translation quality of high-resource languages continues to improve as the corpora expand. These high-resource languages, rich in corpora, include languages such as English, Chinese, Japanese, and French.

However, the translation quality of low-resource languages (such as Tibetan, Uyghur, Burmese, and Urdu in Asia; Irish, Welsh, and Basque in Europe; and Sokoto and Jacque in Africa) remains unsatisfactory. Apart from a few resource-rich languages like English, German, and Chinese, most languages in the world cannot find large-scale bilingual parallel corpora to meet the needs of neural machine translation models. There are over 7,000 languages in the world today, and research and development in machine translation systems are mostly concentrated on languages like English or other languages with large-scale text corpora (around 20 languages). Most of the world's other languages are in urgent need of corresponding language processing tools and corpus resources to meet the computational demands of current deep learning models.

In order to further promote the development of machine translation for low-resource languages, accelerate research progress, enhance technical exchanges, and integrate technology with real-world applications, various influential academic activities have been actively conducted internationally in recent years. One of the most authoritative international competitions in the field of machine translation, the Conference on Machine Translation (WMT), has included translation evaluation tasks from English to low-resource languages in the domain of news translation in recent years. Additionally, the Workshop on Asian Translation (WAT2) focuses specifically on Asian language translation and has been held continuously for six sessions as of 2019. This conference features translation evaluation activities between low-resource Asian languages and English. Therefore, multilingual translation methods are essential for low-resource translation tasks because translation models from multiple languages can help resource-limited languages gain additional knowledge. Moreover, the inclusion of different languages enhances the model's ability to generalize better compared to bilingual translation, thereby improving transfer learning capabilities.

2. Research Background

Currently, China is vigorously promoting the Belt and Road Initiative and building a community with a shared future for mankind. As of the second half of 2021, 172 countries and international organizations have signed over 200 cooperation documents with China to jointly build the Belt and Road. Language communication is an essential guarantee for the realization of the Belt and Road construction, and machine translation is the accelerator and catalyst for achieving language communication. Language serves as humanity's most crucial means of communication and vehicle of information. It forms the foundational element of culture and stands as a distinctive emblem. It is often referred to as the "best key to understanding a country." Therefore, in the future, we should further strengthen the research on machine translation between Chinese and low-resource languages, especially the languages of countries and regions along the Belt and Road. This includes building diversified language data resources and using different methods to comprehensively improve translation quality. Research on machine translation for domestic minority languages has been progressing relatively slowly due to the scarcity of corpus resources and the unique characteristics of the languages. Currently, machine translation research is mainly focused on a few minority languages such as Mongolian, Tibetan, and

Uyghur. Research by the National Language Resources Monitoring and Research Center has found that among the countries along the Belt and Road Initiative proposed by the Chinese government, 9 countries use minority languages or Chinese, which are also used within China, as their national languages or official languages. In the Middle Eastern region, many Arab countries use Arabic, which is similar to Uyghur within China. Nepal and Bhutan use variants of Tibetan. Apart from these Belt and Road countries, in East Asia, South Korea and North Korea are also located near China's northeastern border and use Korean, which is closely related to the Korean language spoken in the Yanbian region of China. Therefore, there is a close connection between the languages of the Belt and Road countries and China's minority languages and Chinese, covering a total of 33 countries and involving 11 languages. Due to the relatively scarce language resources of countries along the Belt and Road, relevant machine translation research is mainly focused on translation between these languages and English, and it largely remains in the statistical machine translation stage. For example, Urdu belongs to the Indo-European language family and is morphologically rich, while Chinese is morphologically sparse, posing challenges for machine translation research between Urdu and Chinese.

Therefore, given the still very limited amount of NMT research between Chinese and other low-resource languages, and the urgent need to further strengthen machine translation research between Chinese and low-resource languages, especially the languages of countries and regions along the 'Belt and Road' initiative, the researchers have chosen to focus on improving the machine translation quality of low-resource languages from countries along the 'Belt and Road' as the central topic of this study.

3. Research Status

Currently, the mainstream approaches to multilingual translation models mainly include three strategies: 1) using different encoders and decoders to represent different languages; 2) using a unified encoder and decoder for translation between all languages; 3) sharing some parameters in the encoder and decoder to represent common features among languages, while others represent language-specific attributes. However, as mentioned earlier, these approaches often face challenges such as complex model structures, increased computational complexity, and barriers to knowledge sharing.

Existing research mainly explores the following three directions: The first approach, in the absence of sufficient bilingual parallel corpora, proposes an innovative data augmentation framework based on a method of semantically related word substitution. Given the limited bilingual parallel corpora, this framework aims to expand the parallel corpora, thereby improving the performance of neural machine translation. The second approach addresses the translation tasks of low-resource languages without parallel corpora. By utilizing only large-scale monolingual data, a new unsupervised neural machine translation framework is proposed. This framework aims to enhance cross-linguistic word embedding performance by mining parallel sentence pairs from monolingual data, thereby improving the performance of unsupervised neural machine translation. The third approach addresses the translation tasks between Chinese and other languages that lack parallel sentence pairs. By utilizing only the parallel sentence pairs between the source language-English and English-target language, a new neural machine translation framework is proposed, which uses English as a pivot language. This framework initializes the translation model through a dual model, and then integrates various pivot-based translation models to construct parallel corpora between the source and target languages, thereby aiming to improve the performance of neural machine translation.

The above three research methods have all improved the translation quality between low-resource languages. Particularly, the third method is the most widely used and has the most significant effect. It employs a strategy known as Pivot Prompting, using English, the most widely used language, as the pivot language for translation between low-resource languages and other languages. This method improves translation quality through the steps of source language to English to target language.

Additionally, the aforementioned Pivot Prompting is actually an enhancement based on prior research on Pivot Translation Techniques. Pivot Prompting and the previous Pivot Translation Technique share similarities as they both employ intermediate pivot points or intermediary steps in the generation model for task execution. However, Pivot Prompting is more suitable for Chat GPT as it can simultaneously recognize the [PIV] language and the [TGT] language, translating the source sentence sequentially into both languages.

Based on the current available data, the third method involving pivot languages for translation shows a high research interest and significant development potential, with ample room for further exploration. Existing research indicates that Pivot Prompting, fundamentally similar to previous Pivot Translation Techniques, is particularly advantageous for Chat GPT applications. For commonly used multilingual sequence-to-sequence translation models, Pivot Translation Techniques typically involve two steps:

1) Inputting the source sentence and translating it into a pivot language.
2) Taking the translated result in the pivot language and translating it into the target language.

In contrast, Chat GPT can recognize both [PIV] (pivot language) and [TGT] (target language) simultaneously, translating the source sentence into both languages sequentially in a single step. Please input a command here: 'Please provide the English translation first and then the Chinese translation for these sentences one by one.'

In summary, the translation quality on languages with abundant data resources continues to improve, but remains unsatisfactory for languages with scarce resources. Machine translation of low-resource languages is currently a major research focus in the field of machine translation, attracting widespread attention both domestically and internationally in recent years. Research on machine translation explores how computers can achieve automatic translation between natural languages, a significant area within artificial intelligence and natural language processing. In recent years, neural machine translation based on deep learning has rapidly developed, replacing traditional statistical machine translation as the new mainstream method in academia and industry.

Considering the distribution of languages involved in the research, the distribution of low-resource languages is highly uneven. Most work is concentrated on European language scenarios, whether as pivot languages or in translation with other low-resource languages. English is undoubtedly the most widely used language. Only a small portion of the research focuses on languages from Asia and other regions, resulting in a significant gap in machine translation research between Chinese and low-resource languages. This presents new challenges for machine translation and application research, making machine translation for scarce-resource languages one of the important research hot spots in the current field of machine translation. It has also attracted widespread attention domestically and internationally in recent years. This study aims to further and more thoroughly research the third method to improve translation quality between low-resource languages and Chinese, delving into the extent to which this method enhances translation performance for low-resource languages.

4. Research Content

This study aims to improve the machine translation quality between low-resource languages of countries along the Belt and Road. It seeks to explore methods to enhance the translation quality of low-resource languages and evaluate their effectiveness. The specific research content includes:

4.1 Research on Improving Translation Quality Based on the Pivot Translation Technique:

This research explores the improvement of machine translation quality from low-resource languages of countries along the Belt and Road to Chinese through the Pivot Translation Technique. The Pivot Translation Technique involves using a third language as a pivot language (usually English) to build a bridge between the source language and the target language, thereby enhancing the machine translation quality between the two languages. The main steps of this method are:

First, input the source language [SRC] into a machine translation engine (such as Baidu Translate, Youdao Translate, Google Translate) and translate it into the pivot language (English) [PIV].

Next, translate the pivot language into the target language [TGT] through the machine translation engine.

Then, compare the results of translating directly from the source language to the target language with the results of using English as the pivot language.

Finally, evaluate the degree of improvement in translation quality based on the results of the first two steps (see Figure 1).

[SRC] → [PIV] → [TGT]

Figure 1

4.2 Research on Improving Translation Quality Based on Pivot Prompting Method:

This research explores the enhancement of machine translation quality from low-resource languages of countries along the Belt and Road to Chinese using the Pivot Prompting method with Chat GPT. Pivot Prompting shares similarities with the previous Pivot Translation Technique—it's a technique that introduces intermediate prompts or intermediary steps in the generation model to perform tasks. However, Pivot Prompting is more suitable for Chat GPT, as it can simultaneously recognize the [PIV] language and the [TGT] language, sequentially translating the source sentences into both languages. The main operational steps of this method are:

Inputting instructions into Chat GPT (e.g., "Please provide the [PIV] translation first and then the [TGT] translation for these sentences one by one").

Chat GPT rapidly generates translations in both the pivot language and the target language, combining the two steps required by Pivot Translation Technique into one.

Evaluating the effectiveness based on the degree of improvement in translation quality for both the pivot language and the target language.

4.3 Evaluation of Translation Quality Improvement:

This study aims to assess the degree of improvement in machine translation quality between low-resource languages of countries along the Belt and Road and Chinese using a combined approach of automatic evaluation and human evaluation methods.

Automatic Evaluation: Automatic evaluation metrics such as COMET, BLEU, and METEOR will be utilized. These metrics offer efficiency, objectivity, standardization, and quantifiability in evaluating machine translation quality. However, they have limitations such as metric specificity, dependency on reference translations, and lack of suitability for specific contexts or singular evaluations.

Human Evaluation: Human evaluation will be conducted using MQM (Multidimensional Quality Metrics). MQM is a comprehensive, open, and customizable quality assessment framework that provides a hierarchical error classification system. It constructs a family of related evaluation metrics based on this framework. The MQM model is systematic, with evaluation metrics (Metrics) occupying a central position. Its usage is flexible; users can either adopt the model's preset metrics for quality assessment or customize them according to their needs.

The evaluation will be conducted across six dimensions:

Accuracy: The correctness of the translation.

Fluency: How naturally and idiomatically the translation reads.

Consistency: The coherence and uniformity of style and terminology throughout the translation.

Grammar and Spelling: The correctness of grammar and spelling.

Coherence: The logical flow and connection of ideas in the translation.

Style and Tone: The appropriateness and fidelity of the translation to the original style and tone.

5. Research Technical Scheme

This study aims to improve the machine translation quality between low-resource languages of countries along the Belt and Road and Chinese. Five official languages of countries along the Belt and Road are selected as low-resource languages for this study: Urdu, Burmese, Kyrgyz, Kazakh, and Tajik, which are the official languages of Pakistan, Myanmar, Kyrgyzstan, Kazakhstan, and Tajikistan, respectively. The specific technical steps are as follows (see Figure 2):

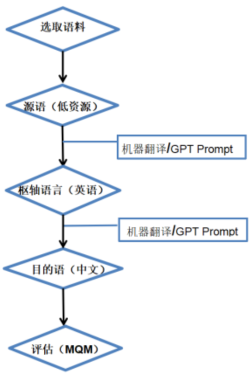

Figure 2

Step 1: Selection of Corpus:

The study uses the Flores-200 multilingual parallel corpus, extracting parallel corpora for the five selected languages. The sampling method is purposive sampling, aiming to select 500 complex sentences (approximately 10,000 words) for each of the five languages as samples for automatic evaluation. From these, 10% (50 sentences) will be selected as samples for human evaluation.

Step 2: Translate Source Language (Low-Resource Languages) to Pivot Language (English):

Researchers will use both the Pivot Translation Technique and the Pivot Prompting method for translation. The Pivot Translation Technique involves inputting the source language [SRC] into machine translation engines (Baidu Translate, Youdao Translate, Google Translate) to translate it into the pivot language (English) [PIV]. The Pivot Prompting method involves inputting the command into Chat GPT (e.g., "Please provide the [PIV] translation first") to generate the pivot language.

Step 3: Translate Pivot Language (English) to Target Language (Chinese):

Researchers will use both the Pivot Translation Technique and the Pivot Prompting method for translation. The Pivot Translation Technique involves using machine translation engines (Baidu Translate, Youdao Translate, Google Translate) to translate the pivot language into the target language [TGT]. The Pivot Prompting method involves inputting the command into Chat GPT (e.g., "then the [TGT] translation for these sentences one by one") to quickly generate the target language.

Step 4: Evaluation of Translation Quality:

The study will use a combination of automatic and human evaluation methods to assess the improvement in machine translation quality between low-resource languages of countries along the Belt and Road and Chinese. Automatic evaluation will employ metrics such as COMET, BLEU, and METEOR.

BLEU is a widely used machine translation evaluation metric that calculates the score by comparing n-gram overlaps between machine translation output and human reference translations. A higher BLEU score indicates that the machine translation output is closer to the reference translation. However, it has some limitations, such as weaker evaluation capability for long sentences and complex structures.

METEOR combines word-level overlap and semantic similarity calculations by evaluating the number of matches, insertions, and deletions in the translation. It also considers word semantic information, providing better judgment capability for synonym and near-synonym substitutions compared to BLEU and NIST.

Automatic evaluation can quickly assess a large number of machine translation outputs, improving evaluation efficiency. Compared to human evaluation, automatic evaluation can significantly reduce time and labor costs. It is based on predefined metrics and algorithms, ensuring objectivity. The evaluation metrics and algorithms used in automatic evaluation are standardized, which means they can be compared and replicated across different datasets and tasks. This helps to promote research progress and the exchange of results in the field of machine translation. Unlike human evaluation, which can be influenced by subjective factors, automatic evaluation provides more objective evaluation criteria. The scores produced by automatic evaluation quantitatively reflect the performance of the machine translation system, providing specific feedback for researchers and developers. This helps to identify and address issues within the system and guides its improvement and optimization.

Automatic evaluation methods can be applied to various types of machine translation systems and tasks, including statistical machine translation, neural machine translation, and translation tasks between different language pairs. In summary, automatic evaluation offers several advantages in the field of machine translation, including efficiency, objectivity, standardization, and quantifiability. However, it also has some drawbacks, such as metric limitations, dependency on reference translations, unsuitability for specific scenarios, and singularity. Therefore, researchers also employed a human evaluation metric—MQM (Multidimensional Quality Metrics). MQM is a comprehensive, open, and customizable quality evaluation framework. Within this framework, it provides a hierarchical error classification system and constructs a family of related evaluation metrics based on this system. The MQM model is systematic, with evaluation metrics (Metric) at its core. The model's usage is quite flexible, allowing users to either directly use the preset metrics for quality evaluation or customize them. This study intends to evaluate translation results from six dimensions: Accuracy, Fluency, Consistency, Grammar and Spelling, Coherence, and Style and Tone.

6. Research Conclusion

The current research is ongoing. Researchers have extracted parallel corpora for five languages from the Flores200 corpus, selecting 500 long and complex sentences for each language. They are using both the Pivot Translation Technique and Pivot Prompting methods to translate the source languages into a pivot language (English) via machine translation engines (Baidu Translate, Youdao Translate, Google Translate) and Chat GPT. Subsequently, the same methods are employed to translate English into Chinese. The quality of the translations is then assessed using both automated evaluations (COMET, BLEU, METEOR) and manual evaluations (MQM).

The study has found that the quality of machine translation is somewhat dependent on the chosen translation methods and tools. Pivot Translation Technique and Pivot Prompting exhibit different effects on translations between different language pairs. Notably, Pivot Prompting, leveraging ChatGPT's powerful natural language processing capabilities, enhances translation fluency and accuracy. While automated evaluation methods offer efficient, objective, and standardized assessments, they have limitations in handling specific scenarios and complex sentences. Therefore, comprehensive evaluation requires combining both automated and manual assessments.

The research data is still undergoing repeated verification, but the emerging trend suggests that using a pivot language yields better translation quality, with higher accuracy and fluency in the translations.

7. Scientific Significance of Research

Researching how to improve machine translation quality from low-resource languages to Chinese within the context of the "Belt and Road Initiative" has significant scientific research implications and socio-economic benefits, as detailed below:

1) Development of Language Technology:

By studying the machine translation from low-resource languages to Chinese, this research can advance the field of language technology. This includes improvements in machine translation models, innovations in cross-language information retrieval technologies, and optimizations of language generation models, bringing new breakthroughs and progress to the field of natural language processing.

2) Linguistic Theory Research:

Machine translation is an interdisciplinary field involving linguistics, computer science, and engineering. Studying the translation issues from low-resource languages to Chinese involves the exploration and validation of linguistic theories. By researching the conversion rules and language structures between different languages, a deeper understanding of linguistic principles can be achieved, promoting the development of linguistic theories.

3) Cross-Cultural Communication:

Enhancing the machine translation quality from low-resource languages to Chinese can promote cross-cultural communication among countries along the "Belt and Road." This helps to strengthen mutual understanding and friendly cooperation between countries, fostering regional cultural prosperity and common development.

From the application of Pivot Prompting in translating low-resource languages to Chinese, there are several benefits:

1) Cross-Language Information Retrieval:

Translation of low-resource languages may lack large-scale parallel corpora, so Pivot Prompting can serve as a solution for cross-language information retrieval. Studying the application of Pivot Prompting in cross-language information retrieval helps explore how to use an intermediate language (such as Chinese) to improve information exchange between low-resource languages.

2) Improvement of Language Generation Models:

In the translation from low-resource languages to Chinese, Pivot Prompting can help improve language generation models, particularly in terms of generation quality and accuracy for rare languages. This contributes to a deeper understanding of semantic relationships between languages and promotes the development of language generation models.

3) Cultural Exchange and Understanding:

By studying the application of Pivot Prompting in translating low-resource languages to Chinese, cultural exchange and understanding between different cultures can be promoted. This is significant for cooperation and communication among countries along the "Belt and Road," helping to strengthen cultural ties and friendly relations.

Therefore, researching how to improve the machine translation quality from low-resource languages to Chinese within the "Belt and Road" initiative can promote regional cooperation and development, with important scientific research significance and socio-economic benefits.

References

[1] Artetxe, M., Labaka, G., Agirre, E., & Cho, K. (2018). Unsupervised Neural Machine Translation. In Proceedings of the International Conference on Learning

Representations (ICLR)2018. arXiv:1710.11041 [cs.CL].
[2] Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural Machine Translation by Jointly Learning to Align and Translate. Proceedings of the International Conference on Learning Representations (ICLR).
[3] Currey, A., Miceli-Barone, A. V., & Heafield, K. (2017). Copied monolingual data improves low-resource neural machine translation. In Proceedings of the Second Conference on Machine Translation. Retrieved from aclanthology.org
[4] Dewangan, S., Alva, S., Joshi, N., & Bhattacharyya, P. (2021). Experience of neural machine translation between Indian languages. Machine Translation, 35(3-4), 71-99.
[5] González-Rubio, J., Navarro-Cerdán, J. R., & Casacuberta, F. (2013). Dimensionality reduction methods for machine translation quality estimation. Machine Translation, 27(3-4), 281-301.
[6] Haque, R., Liu, C.-H., & Way, A. (2021). Recent advances of low-resource neural machine translation. Machine Translation, 35, 451-474.
[7] Kumar, S., Anastasopoulos, A., Wintner, S., & Tsvetkov, Y. (2021). Machine Translation into Low-resource Language Varieties. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.), Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers) (pp. 110-121). Association for Computational Linguistics.
[8] Lavie, A., & Denkowski, M. J. (2009). The METEOR metric for automatic evaluation of machine translation. Machine Translation, 23(3-4), 105-115.
[9] Liu, C.-H., Karakanta, A., Tong, A. N., Aulov, O., Soboroff, I. M., Washington, J., & Zhao, X. (2020). Introduction to the Special Issue on Machine Translation for Low-Resource Languages. Machine Translation, 35(1), 247-249.
[10] Liu, C.-H., Karakanta, A., Tong, A. N., Aulov, O., Soboroff, I. M., Washington, J., & Zhao, X. (2021). Introduction to the second issue on machine translation for low-resource languages. Machine Translation, 35(1), 1-2.