Region-Educational Research and Reviews, 2025,Volume 7, Issue4
DOI:10.12238/rerr.v7i4.3864
ISSN Online:2661-4634
ISSN Print:2661-4626

Research on English Translation System Combining Statistical Machine Translation and Deep Learning

Xiaoli Yu

English College, Zhejiang Yuexiu University

Abstract: With the rapid development of machine translation technology, Statistical Machine Translation (SMT) and Neural Machine Translation (NMT) have become the two dominant translation methods. In recent years, hybrid models combining these two methods have garnered increasing attention due to their ability to integrate the advantages of both approaches, thereby improving translation quality. This paper compares the translation performance of three models (SMT, NMT, and hybrid models), designs a series of experiments, and provides a detailed analysis of the experimental results. The results indicate that the hybrid model outperforms both the SMT and NMT models in terms of BLEU score, TER score, and human evaluation, proving the effectiveness of combining statistical machine translation with deep learning. The research in this paper provides new ideas and methods for the further development of machine translation technology.

Keywords: statistical machine translation, deep learning, neural machine translation, hybrid model

Introduction

With the rapid advancement of information technology and globalization, machine translation (MT) technology has found widespread application across numerous fields, ranging from automatic translation on multilingual websites to facilitating cross-language communication for global enterprises. The importance of reliable and efficient translation systems has thus become more pronounced. Traditional Statistical Machine Translation (SMT) methods rely on large-scale bilingual corpora and statistical techniques to build translation models. By analyzing vast parallel corpora, SMT extracts regularities in vocabulary, syntax, and semantics, enabling automatic language conversion. However, as technology has progressed, SMT methods have increasingly shown limitations, particularly in handling complex sentence structures and long texts. These shortcomings often lead to translations that are inaccurate or sound unnatural, as SMT struggles to account for nuances in context and syntax. In recent years, Neural Machine Translation (NMT) technology has emerged as a game-changer, dramatically enhancing translation quality. Unlike SMT, which is based on statistical principles, NMT uses deep learning techniques to directly map the relationship between the source and target languages through neural networks, allowing for a more nuanced understanding of context. This results in more fluent and accurate translations, especially in capturing the subtleties of language. The successful implementation of NMT marks a significant milestone in the evolution of machine translation technology, offering a more efficient and effective solution for overcoming language barriers in global communication[1].

2. The Importance of Combining Statistical Machine Translation and Deep Learning

2.1 Limitations of statistical machine translation

Statistical machine translation is a translation technique based on probabilistic modeling, which uses a large number of bilingual corpus for model training and adopts statistical methods to generate the most probable translation results.Although SMT has good performance on large-scale datasets, it is weakly dependent on linguistic context, and it cannot deal with the complex syntactic structure and semantic information well.SMT has high requirements on features, and the translation effect depends on the quality of dictionaries and translation rules, and is susceptible to the interference of the above factors[2].

2.2 Advantages and challenges of deep learning translation models

Neural Machine Translation (NMT) is an emerging deep learning technology in recent years, which learns through an end-to-end neural network and is able to directly convert the input source language sequence into the target language sequence. Compared with SMT NMT can automatically learn syntactic and semantic information of data and process long distance dependencies and contextual information. NMT has shown better performance than SMT in many translation tasks. However, NMT requires high computational resources, especially high-performance hardware support during large-scale data training, and model interpretability is not strong[3].

2.3 Advantages of combining SMT and NMT

By combining the strengths of Statistical Machine Translation (SMT) and Neural Machine Translation (NMT) into a hybrid model, it is possible to leverage the statistical characteristics of SMT alongside the contextual understanding capabilities of NMT. This hybrid approach effectively integrates the benefits of both methods, allowing for a more robust translation system. The model can utilize techniques such as weighted averaging and integrated learning to merge the outputs of both SMT and NMT, ensuring that each model contributes its strengths to the final translation.

The hybrid model is particularly adept at handling complex linguistic structures and the rich contextual information often required in translation tasks. Additionally, it capitalizes on the computational efficiency of SMT, which excels in processing large-scale data quickly, while also benefiting from the deeper contextual insights provided by NMT. This balanced combination allows the hybrid model to maintain high translation quality without compromising on speed, thereby addressing the challenges of both translation accuracy and computational efficiency in a variety of language contexts[4].

3.Experimental Design

3.1 Materials and Instruments

The translation dataset used in this experiment is the WMT (Workshop on Machine Translation) dataset, which includes 1000 English-Chinese translation pairs. All experiments were conducted on Linux platform, and the TensorFlow framework was used to implement the NMT model, the Moses tool to implement the SMT model, and the hybrid model was implemented based on the weighted average strategy.

3.2 Setting of experimental and control groups

The experimental group in this study includes three different translation models: Statistical Machine Translation (SMT), Deep Learning-based Neural Machine Translation (NMT), and a hybrid model that combines both SMT and NMT approaches. The purpose of the experimental group is to evaluate the comparative performance of these models by assessing their translation quality, efficiency, and contextual understanding. On the other hand, the control group consists of traditional rule-based machine translation (RBMT) systems, as well as other publicly available machine translation systems, which serve as benchmarks for evaluating the effectiveness of the experimental models. By contrasting the hybrid model and the other models with these established systems, the study aims to determine whether the combination of SMT and NMT offers tangible improvements in translation quality and efficiency, especially in handling more complex and context-sensitive translation tasks.

3.3 Experimental steps

Data preprocessing: clean and segment the data to ensure the data quality.

Model training: use SMT tool to train the translation model, and use deep learning framework to train the NMT model.

Model Integration: Use the weighted average method to combine the translation results of SMT and NMT models into a hybrid model.

Translation Evaluation: Evaluate the translation quality using BLEU, TER and manual evaluation.

Data statistics and analysis: the experimental results are statistically analyzed and tested for significance.

4. Experimental results and analysis

4.1 Comparison of model translation results

Table 1 presents in detail the BLEU scores, TER scores, and scores of human evaluation of SMT, NMT, and hybrid models in the experimental data. The experimental results show that the BLEU scores of the NMT model are significantly better than the scores of the SMT model, and the hybrid model has the optimal performance of each assessment index. Combining the SMT and NMT model characteristics, the hybrid model achieves better results in terms of both accuracy and fluency.

Table 1: Comparison of Translation Performance of Different Models

Model	BLEU Score	TER Score	Human Evaluation Score (1-5)
SMT	0.32	0.35	3.1
NMT	0.45	0.28	4.2
Hybrid Model	0.48	0.26	4.5

From Table 1, it can be seen that the BLEU score of the NMT model is significantly higher than that of the SMT model, which indicates that it performs better in dealing with translation quality, especially when contextual information is taken into account. And the hybrid model not only exceeds the NMT model's score in terms of BLEU score, but also has the best performance in terms of TER score as well as manual evaluation score. By integrating the respective advantages of SMT and NMT, the hybrid model can reduce the errors occurring in the translation process while maintaining a high degree of accuracy, thus improving the overall translation quality level.

4.2 Analysis of Assessment Indicators

The experiment uses BLEU score and TER score as the main criteria for translation quality evaluation. From Table 2, it can be clearly observed that the performance of the NMT model is much better compared with the SMT model, especially in terms of the BLEU score. The scores of the hybrid model perform particularly well, which suggests that the model is able to cope with linguistic complexity more effectively after incorporating the two different approaches. This was further validated by the manual evaluation, where the hybrid model received high ratings for both semantic accuracy and fluency.

Table 2: Comparison of BLEU and TER Scores

Model	BLEU Score	TER Score
SMT	0.32	0.35
NMT	0.45	0.28
Hybrid Model	0.48	0.26

The analysis of Table 2 shows that the BLEU score of the NMT model is significantly higher than that of the SMT model, indicating that it performs better in terms of translation quality, especially in dealing with language fluency and contextual understanding. The performance of the hybrid model is optimal in both metrics, indicating that it can effectively integrate the respective advantages of statistical translation and deep learning, which can not only produce translation results quickly but also ensure higher accuracy and fluency.

4.3 Error Analysis and Exploration of Reasons

Although the hybrid model performs the best in all the evaluation metrics, the translation results still have some errors in some specific cases, especially when dealing with long sentences and polysemous words. Table 3 shows the performance of the various models in translating these complex sentences. the SMT model is more accurate for short sentences, but the translation results are significantly reduced when long sentences are encountered. the NMT model can handle long sentences very well, especially in capturing the syntactic and semantic information in the context, and the hybrid model can circumvent the above problems by combining the strengths of both models in most cases.

Table 3: Performance of Different Models in Translating Complex Sentences

Model	Short Sentence Translation Accuracy	Long Sentence Translation Accuracy	Polysemy Translation Error Rate
SMT	88%	60%	15%
NMT	75%	85%	10%
Hybrid Model	85%	90%	8%

From the data in Table 3, it can be seen that the SMT model has the highest accuracy when dealing with short-sentence translation, but its performance is obviously poor when dealing with long-sentence translation, with an accuracy rate of only 60%.The NMT model shows excellent performance when dealing with long-sentence translation, with an accuracy rate as high as 85%, which is mainly attributed to its excellent contextualization skills. By combining SMT's high-speed processing techniques with NMT's deep insights, the hybrid model is able to efficiently handle long sentences and complex structures, resulting in an accuracy of up to 90% for long-sentence translation. The hybrid model minimizes the translation error rate for polysemous words, indicating that it is able to understand the context of the words better, which in turn gives more accurate translation results[5].

5. Discussion

5.1 Model Advantages and Limitations

The experimental data in this study demonstrate that the hybrid model excels in translation quality, effectively integrating the unique strengths of both Statistical Machine Translation (SMT) and Neural Machine Translation (NMT). This combination allows the hybrid model to outperform individual translation models across various evaluation metrics. By leveraging SMT's high computational efficiency and NMT's advanced contextual understanding, the hybrid model shows significant improvements in translation accuracy, fluency, and semantic consistency[6]. For instance, when handling complex sentence structures, the hybrid model is more effective than traditional SMT in maintaining semantic accuracy, and it also surpasses NMT in translating longer sentences. However, achieving these superior results comes at the cost of substantial computational resource demands. Training on large-scale corpora, in particular, requires significant processing power and time, which may create an efficiency bottleneck when the model is applied in real-time or large-scale translation tasks.

On the other hand, the NMT model, despite its impressive performance, has notable interpretability limitations. NMT is often described as a "black-box" model, meaning the decision-making process behind translations is not transparent to users or researchers. While NMT is capable of producing high-quality translations, it remains difficult to understand the rationale behind specific translation choices, particularly when errors occur. There is a lack of intuitive translation rules or clear diagnostic tools to guide users in optimizing the model. This opacity complicates the process of model debugging and refinement. In comparison, traditional SMT models, although also facing challenges in handling complex language structures, benefit from a more interpretable statistical foundation, making it easier to trace and adjust translation decisions[7]. However, SMT's interpretability does not fully compensate for its inability to manage intricate syntactic and contextual nuances, especially in complex translations.

5.2 Future Research Directions

Future research can improve the performance and adaptability of hybrid models from multiple perspectives. Improving computational efficiency is the main problem faced by current hybrid models. Although the hybrid model can provide superior translation quality, it has a large computational overhead, which may be a performance bottleneck for large-scale translation tasks, especially when real-time translation is required. To address this problem, more effective model optimization methods can be explored in the future, such as using pruning techniques to reduce the number of parameters in the neural network or using more effective training algorithms, such as quantization-based, knowledge distillation as a means to improve the computational efficiency, so that the quality of the translation is guaranteed, while reducing the consumption of computing resources. Parallelization of model training with the help of distributed computing framework may also be an effective means to break through the computational bottleneck[8].

The multimodal hybrid method that combines language model and lexical model is also a direction worthy of in-depth research. The current hybrid model mainly relies on the combination of statistical translation and neural translation, but other types of models such as language models and lexical models are underutilized. The model introduces more linguistic knowledge and grammar rules, which can better cope with certain low-resource languages or a certain domain translation task. Reinforcement learning-based translation models are expected to be the focus of future research. By applying reinforcement learning techniques, the model can make continuous decision-making adjustments and optimization based on the feedback mechanism during translation activities, thus further improving translation quality and fluency. Therefore, future research can focus on the problem of multimodal model fusion to achieve greater breakthroughs in the areas of translation quality, computational efficiency and language understanding [9].

5.3 Limitations of the Experiment

One of the main limitations of this experiment is the small size of the dataset used, which limits the wide applicability of the experimental results. In the field of machine translation, large-scale high-quality datasets are crucial for model training, while only 1,000 pairs of English-to-Chinese translation data were used in this experiment, and although such a scale can be used to evaluate the basic performance of the model, the experimental results may not be representative enough for the performance of large-scale application scenarios. In order to more accurately evaluate the translation effect of the model, future research should consider using larger public datasets, such as WMT or IWSLT, which will help improve the reliability of the experiments and provide more comprehensive data support for translation between different language pairs[10].

The evaluations in the experiments mainly rely on automated evaluation metrics (e.g., BLEU and TER), which, although capable of measuring translation accuracy and fluency to a certain extent, cannot comprehensively reflect all aspects of translation quality. For example, automated assessment systems cannot fully capture contextual understanding and cultural differences in translation, which are important factors in translation quality[11].

6. Conclusion

This study experimentally compares the performance of Statistical Machine Translation (SMT), Neural Machine Translation (NMT), and the hybrid model in terms of translation effectiveness. The experimental results reveal the superior performance of the hybrid model across multiple evaluation criteria. By combining the high efficiency of SMT with the contextual understanding advantage of NMT, the hybrid model significantly improves translation accuracy and fluency, especially when handling long and complex sentences. The integration of both models allows for the strengths of each to complement one another, leading to better overall translation quality[12].

However, despite the impressive results, the hybrid model still faces significant challenges, primarily in terms of high computational overhead and poor interpretability. The need for extensive computational resources during training and the lack of transparency in decision-making processes present obstacles, particularly when it comes to large-scale applications. These issues limit the practicality and scalability of the hybrid model, and they will need to be addressed in future research to fully realize its potential in real-world translation tasks.

About the author:

Xiaoli Yu,（1982.4）female, Han Nationality, born in Shaoxing, Zhejiang province, postgraduate, lecturer, research direction: College English education and English Translation

References

[1] Sveta C D, Kumar A R, Syam B P. Parts of Speech Tagged Phrase-Based Statistical Machine Translation System for English → Mizo Language. SN Computer Science. 2023; 4(6): 15-19.

[2] Liu Pengjuan. Research on Offline Model Training for Large-Scale Distributed Statistical Machine Translation. Automation and Instrumentation. 2023(12): 18-22.

[3] Xu Hong, Huang Xiean. A Review of Pre-Editing and Machine Translation Research (1990—2023). Foreign Language Audio-Visual Teaching. 2023(06): 43-49+112.

[4] Harrydanmu Abdukirim, Hou Yutao, Yao Dengfeng, et al. A Review of Uyghur Language Machine Translation Research. Computer Engineering. 2024; 50(01): 1-16.

[5] He Yuanyuan. Research on Improving the Quality of Cross-Border Tourism Translation through Statistical Machine Translation. Automation and Instrumentation. 2023(09): 201-204.

[6] He Chenghao, Wang Zehui, Teng Junzhe, et al. A Review of Machine Translation. Computer Knowledge and Technology. 2023; 19(21): 31-34.

[7] Xie Gengquan, He Junlin. A Brief Discussion on the Chinese-English Machine Translation Model Based on Kernel Ridge Regression Technology. Journal of Heihe University. 2022; 13(12):181-183.

[8] Lu Chen, Luo Guihua. Comparative Analysis and Statistics of Errors in Machine Translation of Prose. Journal of Luoyang Normal University. 2022; 41(12): 73-78.

[9] Li Zheng. Application Analysis of Statistical Machine Translation Based on Neural Network Language Model. Information and Computer (Theoretical Edition). 2022; 34(22): 109-111.

[10] Yang Yingying. (2022) Application of Machine Translation in RS10 Cloud Platform Products.*[D]*General Research Institute for Mechanical Science, Beijing.

[11] Yang L ,Lin Y .Constructing a University English Translation Course Teaching System Based on POA and MVETC. Open Journal of Modern Linguistics. 2024; 14(06):1140-1158.

[12] Antonios B B ,Andemariam W S ,Asfaha M Y .Pluralistic language policy and multilingual legal texts in Eritrea. Journal of Multilingual and Multicultural Development. 2024; 45(10):4072-4085.