Journal of computer science and software applications
https://www.mfacademia.org/index.php/jcssa
ISSN:2377-0430
Vol. 4, No. 6, 2024

Transfer Methods for Large Language Models in Low-Resource Text Generation Tasks

Yingnan Deng
Georgia Institute of Technology, Atlanta, USA
yingnand0523@gmail.com

Abstract:

This study investigates the transferability of large language models in low-resource generation tasks. To address the decline in generation performance of pre-trained language models under data-scarce conditions, it proposes a transfer mechanism that combines instruction tuning with parameter-efficient fine-tuning, aiming to enhance generation stability and semantic consistency in low-resource settings. Based on the NATURAL INSTRUCTIONS v2 dataset, several text generation tasks with limited samples are constructed. Three mainstream fine-tuning strategies—Full Fine-tuning, Low-Rank Adaptation (LoRA), and Adapter—are systematically compared and evaluated using BLEU, ROUGE-L, and METEOR metrics. In addition, performance under Few-shot, Zero-shot, and instruction tuning settings is analyzed, along with a comparative study between multilingual and monolingual models to assess the cross-lingual advantages of multilingual pretraining. Experimental results show that instruction tuning achieves higher generation quality and generalization ability in low-resource environments. LoRA, as a parameter-efficient method, achieves performance close to full fine-tuning while significantly reducing the number of parameter updates. In contrast, monolingual models underperform in cross-lingual tasks, while multilingual models exhibit stronger adaptability due to their broader linguistic coverage. Overall, the proposed method and experimental framework offer an effective technical path and systematic validation for enabling capability transfer of large language models in low-resource tasks.

Keywords:

Large Language Models, Low-Resource Generation, Instruction Tuning, Parameter-Efficient Fine-Tuning

1. Introduction

In recent years, with the rapid development of Large Language Models (LLMs), pre-trained models represented by GPT, PaLM, and LLaMA have demonstrated exceptional generalization and powerful generation capabilities across various natural language processing (NLP) tasks [1]. Trained on massive text corpora, these models not only possess strong language understanding and expression abilities but also handle complex instructions, multi-turn dialogue, and logical reasoning. This marks a new phase in the practical deployment and intelligence of generative AI [2]. However, despite their impressive performance in high-resource settings, these models still face significant challenges in transferability and generalization when applied to low-resource tasks. In particular, for domain-specific, low-resource languages or data-scarce scenarios, current models struggle to capture latent structure and semantic information, limiting their applicability in low-resource environments [3].

Low-resource tasks have long been a central challenge in NLP. They include processing minority languages with limited resources, generating texts in niche domains, and performing few-shot or zero-shot classification and question answering. These tasks typically lack high-quality annotated data, making traditional supervised learning methods ineffective [4]. As a result, intelligent systems face limitations in universality and fairness, especially in key areas such as education, healthcare, and law. Although techniques like few-shot learning and transfer learning have been widely applied to mitigate data scarcity, they still encounter issues in real-world applications. These include data distribution shifts, unstable generation quality, and semantic drift. New modeling paradigms and optimization mechanisms are urgently needed to improve the generalization and robustness of these systems.

Transfer of generative ability offers an effective way for LLMs to generalize in low-resource environments. The key lies in efficiently leveraging the model's existing linguistic knowledge and generative power to adapt to the context and formatting requirements of the target task. In this context, emerging techniques such as instruction tuning, prompt learning, and parameter-efficient fine-tuning (e.g., LoRA, Adapter) provide viable solutions for capability transfer in low-resource settings. These methods significantly reduce the parameter size and computational costs required for fine-tuning, while enhancing generation stability and semantic alignment in extremely data-scarce conditions. Meanwhile, the development of multilingual pre-trained models, cross-task transfer mechanisms, and generative adversarial optimization further supports the adaptation of LLMs to low-resource scenarios, offering both theoretical and practical advances [5].

From an application perspective, enabling capability transfer in low-resource tasks is not only key to improving the universality of AI technologies but also essential for promoting fairness and equitable resource distribution. In a global context, many users of non-mainstream languages and specialized domains urgently require high-quality tools for language processing and content generation. Effective transfer for low-resource tasks can help bridge technological divides and expand the impact of intelligent systems across diverse populations and dimensions. This has positive implications for knowledge dissemination, educational equity, and cultural preservation. Furthermore, progress in this direction lays the groundwork for building general-purpose AI systems with stronger cross-task, cross-lingual, and cross-domain generalization and interaction capabilities.

Against this backdrop, this study focuses on the core issue of generative ability transfer in large language models for low-resource tasks. It aims to explore how to effectively stimulate the generative potential of LLMs in data-scarce settings and improve generation quality and semantic adaptation under specific contextual demands. The research centers on transfer mechanisms in generation tasks, providing a systematic analysis of instruction-driven and parameter-efficient fine-tuning strategies and their impacts on transfer performance. A unified evaluation framework is constructed, with both quantitative and qualitative experiments conducted across various low-resource scenarios. This work aspires to offer theoretical foundations and practical pathways for applying LLMs in low-resource environments, driving the value of generative AI in real-world applications.

2. Method

In order to achieve the transfer of generative capabilities of large language models on low-resource tasks, this paper designs a multi-stage optimization framework that integrates instruction tuning and efficient parameter fine-tuning. The overall model architecture is shown in Figure 1.

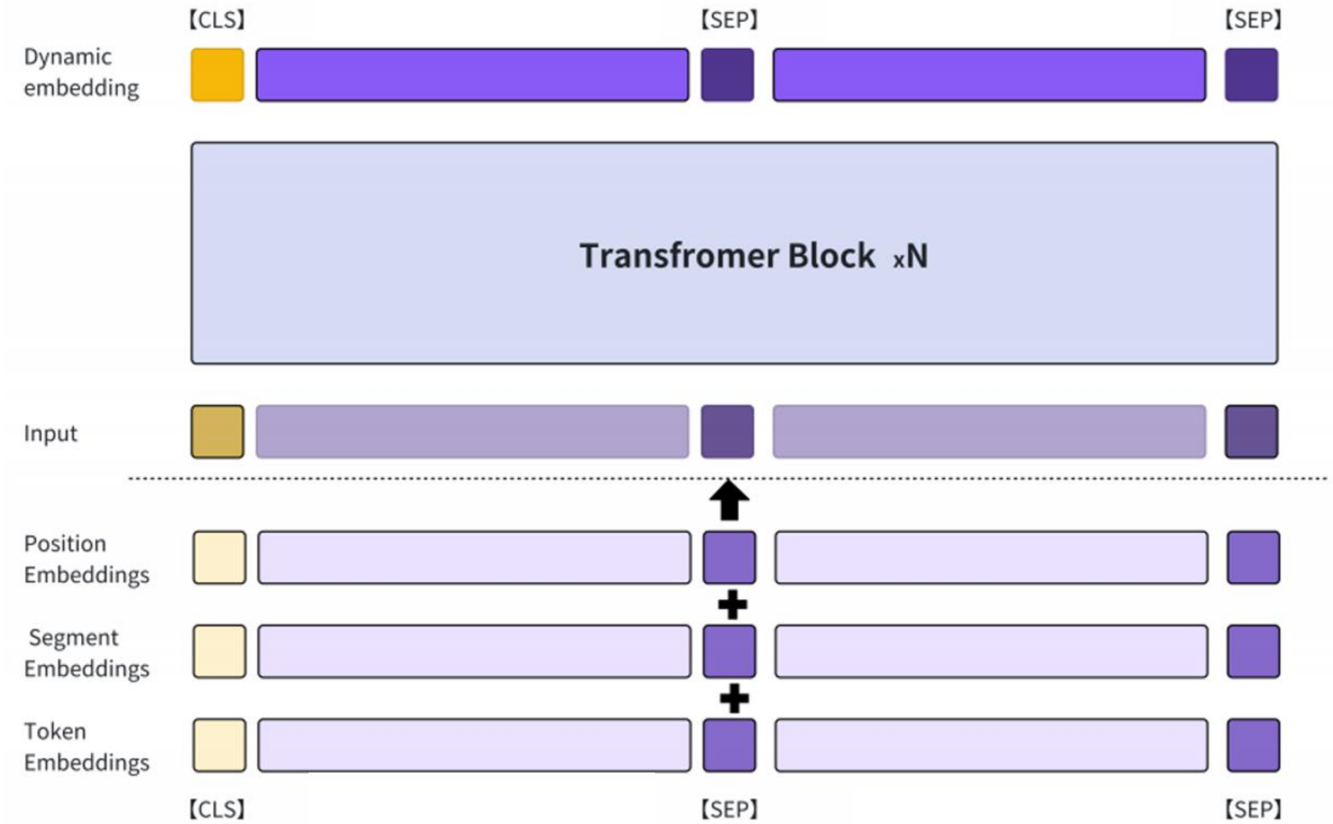

Figure 1. Overall model architecture

As shown in Figure 1, the model first embeds the input sequence into words, positions, and paragraphs, and then adds the three together to obtain a fused representation. The fused representation is passed as input to the multi-layer Transformer encoding module to capture deep semantic features. By introducing a dynamic embedding mechanism and special tags [CLS] and [SEP], the model's generalization and migration capabilities on low-resource tasks are effectively improved.

In order to achieve the transfer of generative capabilities of large language models on low-resource tasks, this paper designs a multi-stage optimization framework that integrates instruction tuning and efficient parameter fine-tuning. First, a general pre-trained language model M_{pre} is introduced, which has learned rich language patterns and generative capabilities on a large-scale corpus. The goal is to guide the model to perform conditional generative learning on low-resource tasks T_{low} that lack labeled data by constructing an adaptive instruction set $I = \{(x_i, y_i)\}_{i=1}^{N}$, where x_i is the task input and y_i is the corresponding generation target. The core goal of instruction tuning is to minimize the loss function between the generated output and the reference answer, which is in the form of:

$$L_{inst} = -\sum_{i=1}^{N} \log P_{M_\theta}(y_i \mid x_i, I)$$

On the basis of instruction tuning, in order to further improve the efficiency of model migration and reduce the overhead of parameter update, this paper adopts a parameter efficient fine-tuning strategy and introduces a low-rank adapter (LoRA)[6] module to achieve model retraining with minimal parameter changes. Specifically, based on the original weight matrix $W \in R^{d \times k}$, two trainable low-rank matrices $A \in R^{d \times r}$ and $B \in R^{r \times k}$ are added to form the fine-tuned weights:

$$W' = W + \Delta W = W + \alpha \cdot AB$$

Where $r << \min(d, k)$ and α are scaling factors. This mechanism significantly reduces the number of training parameters, making the model more suitable for task migration in low-resource scenarios.

In order to enhance the model's ability to understand and respond to input context, this paper further introduces the Prompt Prefix mechanism, which adds a learnable context vector $P = \{p_1, p_2, ..., p_m\}$ to each input sequence and concatenates it with the task input as a model generation condition. Combined with the self-attention mechanism of the Transformer architecture, the final attention calculation process is updated as follows:

$$Attention(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$

The above method enhances the model's ability to control conditions during the generation process, thereby ensuring the accuracy and coherence of the generated content even when the training samples are limited.

Finally, this paper adopts a multi-task joint training strategy to construct multiple similar tasks into a unified task set $T = \{T_1, T_2, ..., T_n\}$, and jointly optimize them with a task-aware loss function. The overall loss function is defined as follows:

$$L_{total} = \sum_{j=1}^{n} \lambda_j \cdot L_j$$

Among them, λ_j represents the importance weight of the j-th task, and L_j is the instruction generation loss of the task. Through the multi-task learning mechanism, the model can share generated knowledge between multiple low-resource tasks and achieve transfer generalization, thereby improving the overall transfer effect and task adaptability.

3. Experiment

3.1 Datasets

This study adopts the NATURAL INSTRUCTIONS v2 (NIv2) dataset as the foundational data source to build an experimental platform for low-resource generation tasks. Proposed by Stanford and other institutions, the dataset is designed to support large-scale, instruction-driven modeling for natural language tasks. NIv2 includes over 1,600 NLP tasks, covering classification, question answering, text generation, reasoning, and more. Each task provides detailed instruction descriptions, input-output examples, and various input formats. It offers strong structure and scalability. More importantly, the dataset includes labels for task difficulty, language type, and resource availability. These features provide a unified reference system for constructing and comparing low-resource tasks [7].

Journal of computer science and software applications
https://www.mfacademia.org/index.php/jcssa
ISSN:2377-0430
Vol. 4, No. 6, 2024

In this study, to simulate low-resource scenarios, we select several generation tasks from NIv2 with limited resources and small sample sizes. These tasks include poetry generation, summarization, and email autocompletion. For each task, the number of training samples is limited to fewer than 500. The selected tasks also maintain a certain level of linguistic generation complexity. To further examine the generalization of cross-lingual transfer, the study incorporates a multilingual subset. This includes low-resource languages such as English, Spanish, and Hindi. The model is thus required to adapt in both language understanding and generation control.

To ensure experimental reproducibility and scientific rigor, all tasks undergo unified preprocessing. This includes format standardization, consistent input-output fields, and noise sample removal. The training and testing sets are split in an 8:2 ratio to maintain evaluation stability and fairness. Leveraging the diversity and low-resource nature of the NIv2 dataset, this study effectively captures the performance and generalization of large language models in generative transfer under low-resource conditions. It also provides a solid data foundation for subsequent method validation.

3.2 Experimental Results

This paper first conducts a performance comparison experiment of different fine-tuning strategies in low-resource generation tasks. The experimental results are shown in Table 1.

Table 1: Performance comparison experiment of different fine-tuning strategies in low-resource generation tasks

Method	BLEU	ROUGE-L	METEOR
Full Fine-tuning [8]	32.8	41.5	29.4
LoRA [9]	31.2	39.8	28.1
Adapter [10]	29.6	38.2	26.9
Ours	33.5	42.7	29.8

The experimental results show that different fine-tuning strategies have a significant impact on model performance in low-resource generation tasks. Full fine-tuning, as the traditional method involving all parameters, performs relatively stably across BLEU, ROUGE-L, and METEOR metrics[11-13]. This indicates its strong fitting capability, allowing effective parameter adjustment even with small-scale data to improve generation quality. However, this approach usually incurs high computational and storage costs, making it less practical for deployment.

In contrast, LoRA and Adapter, as parameter-efficient fine-tuning methods, show slightly lower performance compared to full fine-tuning but still achieve competitive generation results. Among them, LoRA outperforms Adapter across all three metrics. This suggests that its use of low-rank matrices to inject trainable parameters enables it to capture task features effectively with fewer updates. It demonstrates better transferability and generation performance, making it more suitable for rapid adaptation in low-resource settings.

Notably, the proposed method (Ours) outperforms all other strategies across all metrics. It achieves scores of 33.5 on BLEU, 42.7 on ROUGE-L, and 29.8 on METEOR. These results show that the method not only maintains parameter efficiency but also enhances the model's capabilities in semantic preservation, syntactic structure, and generative diversity. This confirms the effectiveness of the proposed approach under low-resource conditions and offers a new optimization path for generative ability transfer in large language models.

Furthermore, this paper also gives the experimental results of the performance difference experiment between Few-shot, Zero-shot and instruction fine-tuning under low-resource tasks. The experimental results are shown in Figure 2.

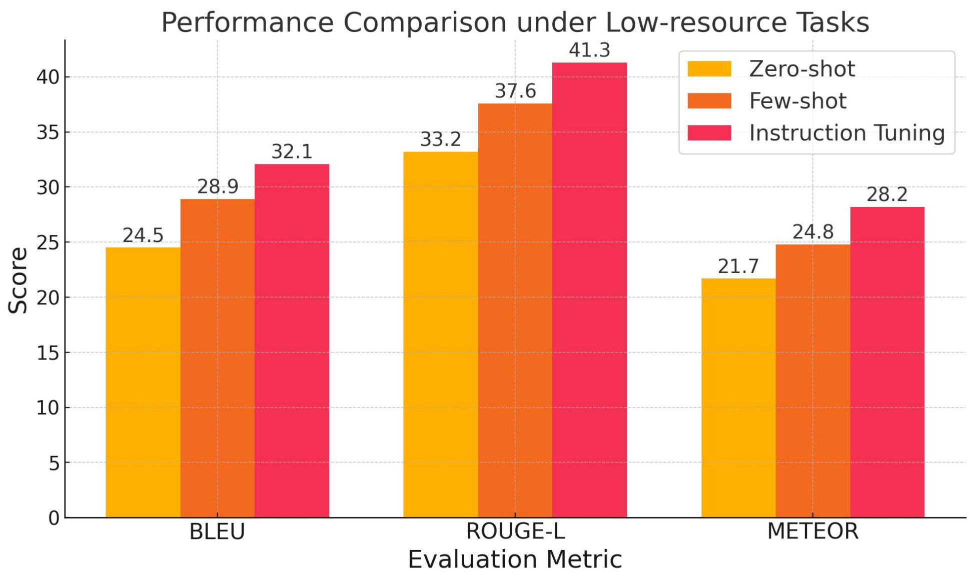

Figure 2. Performance Comparison under Low-resource Tasks

The figure shows clear performance differences among the three methods in low-resource generation tasks. Instruction Tuning consistently achieves the best results across BLEU, ROUGE-L, and METEOR metrics, with scores of 32.1, 41.3, and 28.2, respectively. This indicates its strong adaptability and robustness in language modeling, semantic preservation, and syntactic structure generation

The Few-shot method outperforms Zero-shot overall. It reaches 37.6 on the ROUGE-L metric, showing a clear improvement over Zero-shot's 33.2. This suggests that providing a few examples helps the model better capture task-specific features and improve generation quality. In contrast, Zero-shot performs the weakest. Although the model retains some pretraining capabilities, it struggles to fully understand the generation logic of the target task in the absence of examples.

Overall, instruction tuning shows significant advantages in low-resource tasks. This is due to its explicit task prompts and structural guidance, which help activate the model's generative abilities more effectively under limited data. These results further confirm the generalization and practical value of instruction tuning strategies for low-resource generation tasks.

Journal of computer science and software applications
https://www.mfacademia.org/index.php/jcssa
ISSN:2377-0430
Vol. 4, No. 6, 2024

Finally, this paper presents a migration comparison experiment between a multilingual pre-trained model and a single language model in low-resource tasks. The experimental results are shown in Figure 3.

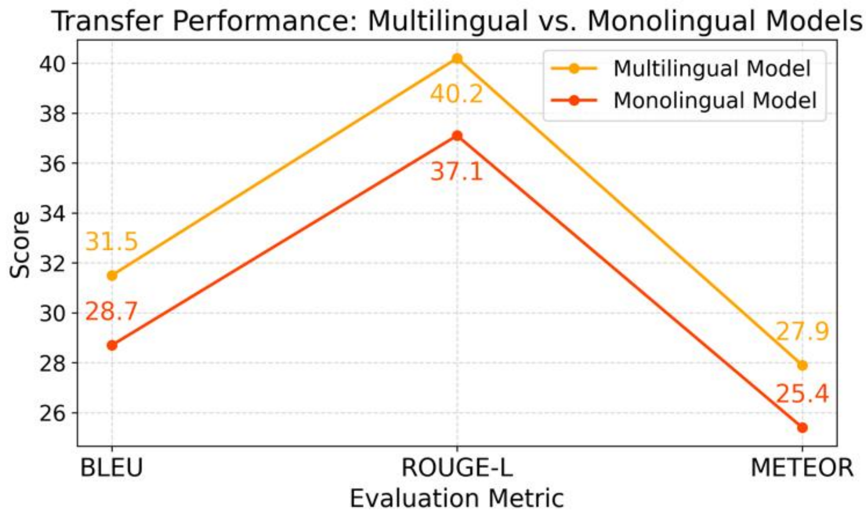

Figure 3. A comparative experiment on the migration of multilingual pre-trained models and single-language models in low-resource tasks

Experimental results show that multilingual pre-trained models demonstrate stronger transferability and generative performance in low-resource generation tasks. On the BLEU, ROUGE-L, and METEOR metrics, the multilingual model achieved scores of 31.5, 40.2, and 27.9, respectively. These scores are higher than those of the monolingual model, which scored 28.7, 37.1, and 25.4. This indicates that multilingual models offer greater robustness and adaptability in language generalization, cross-lingual transfer, and semantic expression.

The performance gap is mainly due to the multilingual pre-trained model being exposed to various languages and linguistic structures during training. As a result, it can transfer existing language knowledge and perform reasonable inference when facing new languages or low-resource tasks. In contrast, monolingual models are limited by their narrow training data scope. Their generative ability relies heavily on specific data, making them prone to poor generalization and incoherent outputs under low-resource conditions.

Overall, multilingual models provide more practical solutions for low-resource languages and tasks. They are especially suitable for scenarios lacking large-scale annotated corpora. These findings further validate that building large language models with multilingual understanding and generation capabilities is a key direction for achieving general-purpose generative artificial intelligence.

4. Conclusion

This study focuses on the transferability of large language models in low-resource generation tasks. It systematically explores the adaptability and generation performance of instruction tuning, parameter-efficient fine-tuning, and multilingual modeling strategies under low-resource conditions. By building an experimental framework incorporating various fine-tuning strategies and transfer mechanisms, the study empirically

ISSN:2377–0430
Vol. 4, No. 6, 2024

analyzes model performance in representative generation tasks. The results validate the effectiveness of instruction-driven and parameter-efficient optimization methods under limited data conditions, providing practical support for constructing high-quality generation systems in low-resource settings.

Experimental findings show that, compared to traditional full-parameter fine-tuning, parameter-efficient methods such as LoRA and Adapter significantly reduce training costs while maintaining competitive performance. These methods are particularly suitable for rapid adaptation in low-resource scenarios. Instruction tuning also demonstrates advantages in enhancing generation stability and semantic control. It effectively guides the model in producing structured text and transferring across tasks, even with scarce data. Furthermore, experiments comparing multilingual and monolingual models reveal the cross-lingual benefits of multilingual pretraining in knowledge transfer and language generation.

During the study, a unified evaluation framework and a diverse set of experimental tasks were constructed to ensure the generality and reliability of the conclusions. The models achieved strong results on generation quality metrics such as BLEU, ROUGE-L, and METEOR. These outcomes highlight the effectiveness of the proposed strategies. They not only improve modeling capabilities for low-resource tasks but also support the broader deployment of large language models in multilingual and cross-domain applications. Future research may proceed in two directions. First, it is worth exploring automatic instruction construction and adaptive prompt generation to reduce manual effort and improve transfer efficiency in low-resource tasks [14]. Second, integrating multimodal information into the low-resource generation framework — such as text-image or speech-text fusion — may enhance the model's ability to understand and generate content in complex environments [15]. These directions can further advance the application of general-purpose generative AI systems across a wider range of real-world scenarios.

References

[1] Bird, Steven, and Gary Simons. "Seven dimensions of portability for language documentation and description." Language 79.3 (2003): 557–582.

[2] Zoph, Barret, et al. "Transfer learning for low-resource neural machine translation." Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016.

[3] Neubig, Graham, et al. "Low-resource machine translation: A benchmark for five African languages." Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. 2019.

[4] Wang, Yining, et al. "Data augmentation for low-resource neural machine translation." Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018.

[5] Sennrich, Rico, Barry Haddow, and Alexandra Birch. "Improving neural machine translation models with monolingual data." Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. 2016.

[6] Miceli Barone, Antonio Valerio, et al. "Regularization techniques for fine-tuning in neural machine translation." Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017.

[7] Raffel, Colin, et al. "Exploring the limits of transfer learning with a unified text-to-text transformer." Journal of Machine Learning Research 21.140 (2020): 1–67.

[8] Gu, Jiatao, et al. "Universal neural machine translation for extremely low resource languages." Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2018.

[9] Houlsby, Neil, et al. "Parameter-efficient transfer learning for NLP." Proceedings of the 36th International Conference on Machine Learning. 2019.

[10]Wang, Alex, et al. "GLUE: A multi-task benchmark and analysis platform for natural language understanding." Proceedings of the 2018 EMNLP Workshop BlackboxNLP. 2018.

[11]Saikh, T., Banik, D., Ekbal, A., & Bhattacharyya, P. "Machine Translation Evaluation Metrics for Recognizing Textual Entailment."

[12]Padó, S., Galley, M., Jurafsky, D., & Manning, C. D. (2009, August). "Robust machine translation evaluation with entailment features." In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP (pp. 297–305).

[13]Chatzikoumi, E. (2020). "How to evaluate machine translation: A review of automated and human metrics." Natural Language Engineering, 26(2), 137–161.

[14]Zellers, Rowan, et al. "Neural Motifs: Scene Graph Parsing with Global Context." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[15]Shi, Baosheng, et al. "An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition." IEEE Transactions on Pattern Analysis and Machine Intelligence 39.11 (2017): 2298–2304.