Multimodal Intermediate-Task Training Effects on Zero-Shot Cross-Lingual Transfer Performance in Language Models

Assignee Research

July 25, 2026

Abstract

In this work we focus on transferring supervision signals of natural language generation (NLG) tasks between multiple languages. We propose to pretrain the encoder and the decoder of a sequence-to-sequence model under both monolingual and cross-lingual settings. The pre-training objective encourages the model to represent different languages in the shared space, so that we can conduct zero-shot cross-lingual transfer. After the pre-training procedure, we use monolingual data to fine-tune the pre-trained model on downstream NLG tasks. Then the sequence-to-sequence model trained in a single lang

1 Introduction

This paper examines: Cross-Lingual Natural Language Generation via Pre-Training. Research question: What is the impact of multimodal intermediate-task training on zero-shot cross-lingual transfer performance for language models when evaluated on XNAT, compared to monolingual text-only intermediate training?.

2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.5/10.

3 Results

4 papers retrieved. 13 claims extracted; 10 independently verified. Quality review score: 7.5/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
XNLG shows positive performance compared with the machine-translation-based pipeline model in zero-shot cross-lingual se	✓	0.21
Shen et al. (2018) and Duan et al. (2019) use translated documents or summaries as pseudo training data for cross-lingua	✓	0.34
Junnan et al. (2019) incorporate monolingual summarization and machine translation to improve cross-lingual summarizatio	✓	0.27
Kumar et al. (2019) use training data annotated in multiple languages to jointly train a sequence-to-sequence model for	✓	0.30
Pre-training BERT on a corpus of multiple languages shows a surprising ability to produce cross-lingual representations	✓	0.26
Lample and Conneau (2019) extend mask language modeling pre-training to cross-lingual settings, showing significant impr	✓	0.19
XNLG pre-trains both encoder and decoder for cross-lingual generation tasks.	×	0.11
Arteetxe and Schwenk (2018) use the sequence encoder of the multilingual translation model to produce cross-lingual sent	✓	0.19
It is difficult to control the target language by directly fine-tuning the pre-trained translation model on downstream N	×	0.14
XNLG is evaluated on two cross-lingual NLG downstream tasks: cross-lingual question generation and cross-lingual abstrac	✓	0.16
XNLG uses a pre-trained model with a 10-layer encoder and a 6-layer decoder.	✓	0.16
XNLG uses 1024 hidden units, 8 attention heads, and GELU activations for every Transformer layer.	✓	0.20
XNLG uses the 15-language pre-trained XLM to initialize the parameters of the English model.	×	0.14

References

- `http://arxiv.org/abs/2505.18673v1`

- http://arxiv.org/abs/2310.10378v5
- http://arxiv.org/abs/1909.10481v3