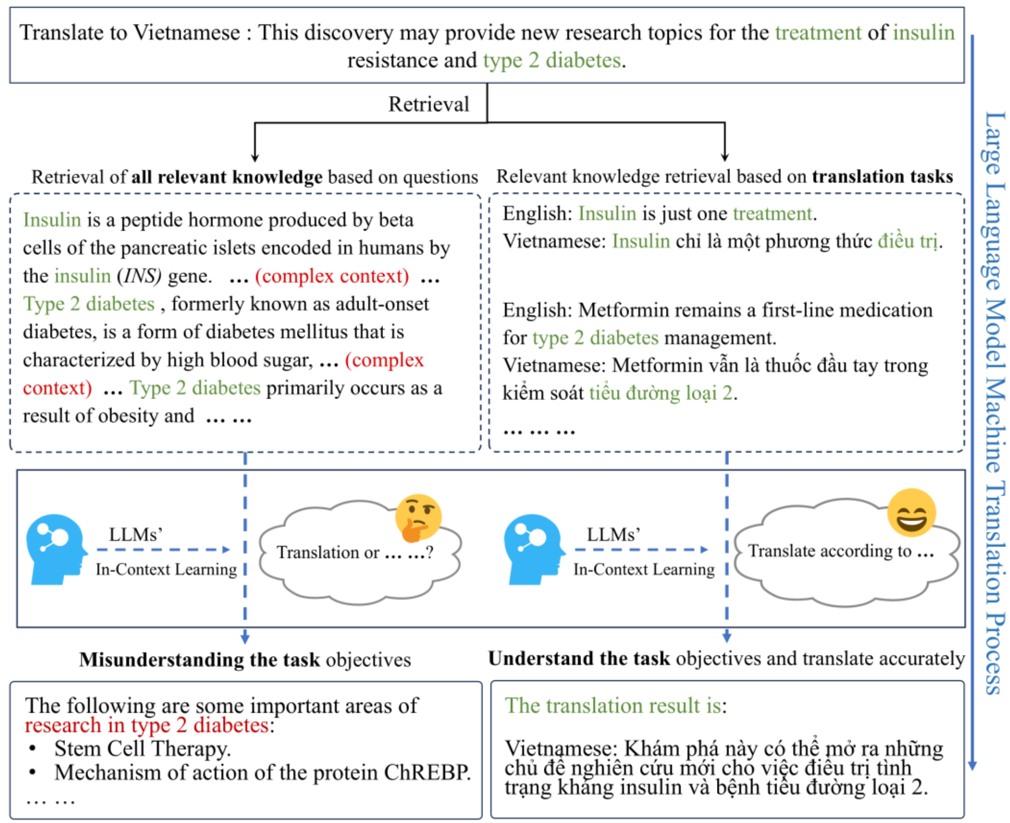
Bridging the Gap: Adapting LLMs for Southeast Asian Low-Resource Machine Translation via Hierarchical Dynamic Retrieval and Matching

Zirui GUO, Hua LAI, Ying LI, Zhengtao YU, Shengxiang GAO, Yuxin HUANG, Cunli MAO

Frontiers of Computer Science, DOI: 10.1007/s11704-026-51767-9

Problems & Ideas

- Problems of retrieval-augmented LLM approaches for low-resource machine translation :
  - Inadequate utilization of retrieved examples in LLM prompting.
  - Lack of multi-granularity matching between source sentences and retrieved contexts.
- Ideas: The key idea is to introduce a hierarchical dynamic retrieval and matching framework tailored for low-resource machine translation.



- Joint matching with symbolic and semantic signals. Improve translation-oriented relevance rather than pure semantic similarity.
- Reordering-based prompt construction for LLMs. Construct highly relevant and coherent bilingual prompts for LLM-based translation

Conventional retrieval methods struggle to find semantically aligned sentences within sparse parallel corpora. While RAG substantially improves LLM performance in general tasks, its effectiveness in low-resource machine translation still depends on robust retrieval strategies.

Main Contributions

- Contributions:
  – A retrieval-augmented framework is introduced to enhance LLM-based low-resource machine translation through dynamic and translation-aware prompting;
  – Competitive translation performance is achieved without any LLM parameter updates, rivaling mainstream NMT systems;
  – Key retrieval and context configurations are analyzed, revealing language-specific optimal settings for Southeast Asian low-resource languages.

Models	ZH-VI		ZH-MY		ZH-ID		ZH-MS	
	BLEU	COMET	BLEU	COMET	BLEU	COMET	BLEU	COMET
NLLB 200-Distilled	28.68	78.19	22.38	71.93	16.33	77.12	13.80	74.86
M2M 100-418M	41.15	83.29	5.56	28.15	24.76	84.19	44.50	83.19
Qwen 2.5-7B	41.17	75.02	8.53	44.58	17.47	79.76	16.02	70.31
+5-Shots	32.82	81.54	12.52	48.51	16.88	83.55	15.98	78.10
+LoRA	37.75	76.50	11.13	47.19	16.87	82.43	12.94	76.13
+Navie RAG	41.82	77.56	14.21	45.43	18.71	82.92	16.23	79.39
+DUAL-REFLECT	36.93	77.39	14.79	47.67	17.79	81.58	12.51	77.42
+Our Method	42.03	81.64	32.02	51.17	18.60	81.39	16.73	79.93
Llama 3.1-8B	41.38	70.71	18.14	45.46	10.58	79.37	26.23	73.37
+5-Shots	44.90	81.46	23.95	53.51	19.73	84.79	16.41	80.85
+LoRA	45.31	79.15	4.64	44.32	20.78	86.25	16.59	83.10
+Navie RAG	43.22	71.63	19.12	44.39	18.65	82.19	21.93	79.34
+DUAL-REFLECT	46.78	81.43	23.59	51.02	11.68	83.49	27.18	79.38
+Our Method	47.44	79.46	24.06	55.91	17.79	81.89	27.35	79.68

Partial machine translation results from the Flores-200 devtest dataset based on LLM and parallel sentence retrieval trees.