BEFT: Bias-Efficient Fine-Tuning of Language Models in Low-Data Regimes

Baichuan Huang[1], Ananth Balashankar[2], Amir Aminifar[1]

[1]Department of Electrical and Information Technology, Lund University, Sweden

[2]Google DeepMind, USA

Motivation

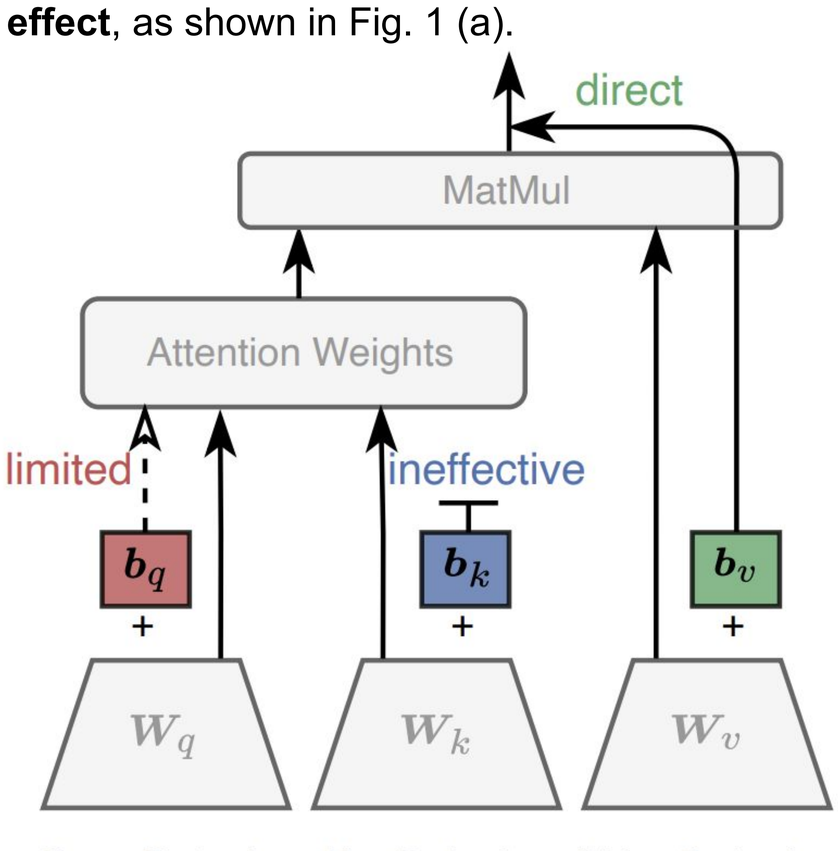
Fine-tuning pre-trained large language models (LLMs) for downstream tasks has gained a lot of attention over the past few years. Parameter-efficient fine-tuning (PEFT) methods have been widely studied to reduce the fine-tuning overheads [1]. Among these PEFT techniques, bias-only fine-tuning—which involves updating only the bias terms of the LLMs—provides the potential for unprecedented parameter efficiency and competitive downstream accuracy especially In low-data regimes [2]. Despite the advantages of bias-only fine-tuning, the relationship between fine-tuning different bias terms and downstream performance remains largely unclear.

Method & Findings

In this paper, we investigate the link between fine-tuning different bias terms (i.e., b_q, b_k, and b_v in the query, key, or value projections) and downstream performance. Specifically, we augment the magnitude changes with angular changes to study the link between fine-tuning different bias terms and downstream performance. **We observe that finetuning b_v generally leads to higher downstream performance in low-data regimes, compared to b_q and b_k.**

We also analyze the potential expressive power of the bias terms query b_q, key b_k, and value b_v for scaled dot-product attention. We find that fine-tuning b_v in low-data regimes is **sufficient**; b_k has **no effect** on improved expressiveness, whereas b_q has a **limited effect**, as shown in Fig. 1 (a).

(a)



(b)

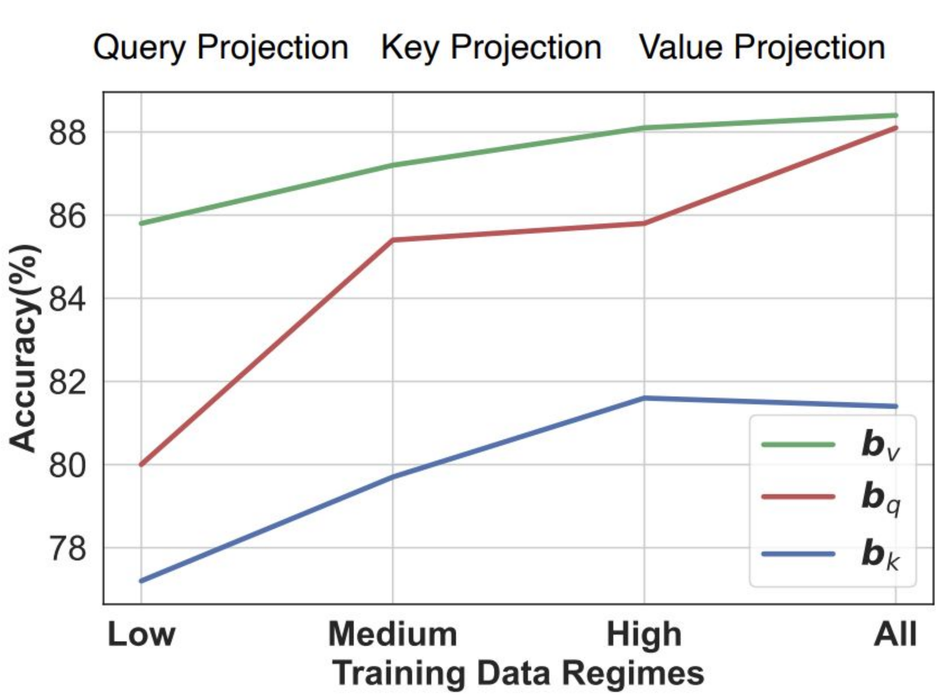


Fig. 1: Expressiveness of b_q, b_k, and b_v.

Empirical Evaluation

Importance Ranking and Downstream Performance

As shown in Fig. 2 and Fig. 1 (b), we expect higher-ranked bias terms to achieve higher accuracy. Mag+Angle (ours) shows a precise and dynamic link between bias-term rankings and downstream performance across diverse data regimes.

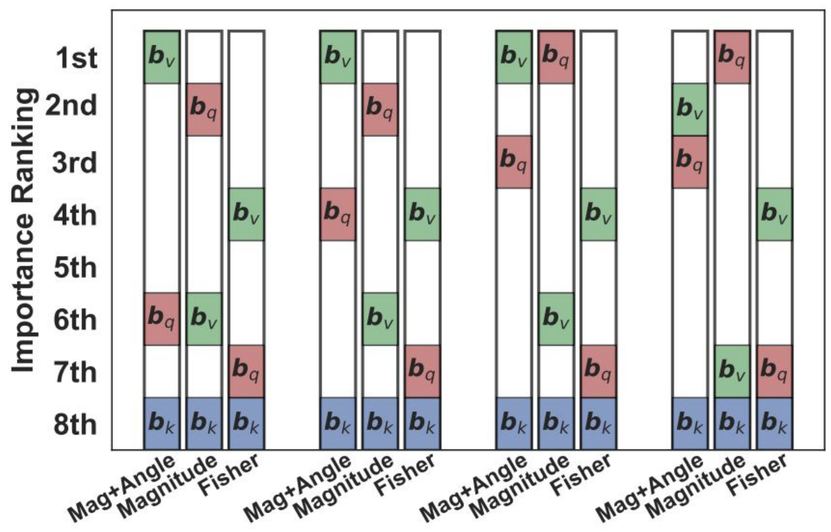


Fig. 2: Fine-tuning b_q, b_k, and b_v ($BERT_{BASE}$ on SST-2).

BEFT can be Combined with Other PEFT Methods

Bias Term	BEFT	+LoRA	+VeRA	+DoRA
b_v	**85.8%**	**85.0%**	**82.2%**	**86.0%**
b_q	80.0%	82.6%	81.8%	83.5%
b_k	77.2%	76.9%	81.7%	76.9%

Table 1: b_v surpasses b_q, and b_k in downstream performance ($BERT_{BASE}$ on SST-2).

Extension to Bias-Free LLMs

Bias Term	Adding b_v	Adding b_q	Adding b_k
LLaMA2-7B	**94.9%**	90.0%	68.0%
DeepSeek-Coder-Base-1.3B	**76.9%**	67.4%	60.3%
GPT-J-6B	**92.8%**	88.3%	63.8%

Table 1: Our key finding still holds (SST-2).

(refer to our paper for more results)

References

[1] Ding, Ning, et al. Parameter-efficient fine-tuning of large-scale pre-trained language models. Nature machine intelligence, 2023.

[2] Zaken, Elad Ben, et al. Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language-models. ACL, 2022.

ACL 2026

WASP Wallenberg AI, Autonomous Systems and Software Program

Swedish Research Council

Swedish Foundation for Strategic Research

ELLIIT Excellence Center at Linköping – Lund in Information Technology

NAISS

huggingface/peft/beft