Dataset	Editor	LLAMA2-7B			LLAMA-7B		
		Reliability	Generality	Locality	Reliability	Generality	Locality
ZsRE	Original model	43.70	43.17	/	43.29	42.85	/
	LoRA	43.10	42.20	70.83	46.93	45.87	75.86
	F-Learning$_{\text{LoRA}}$	**46.91**	**46.21**	**72.50**	**47.56**	**46.90**	**76.09**
	FT-c	49.02	46.96	67.37	47.33	45.51	68.14
	Full-FT	81.02	74.67	70.51	70.52	66.69	65.26
	ROME	43.67	42.66	**93.14**	43.45	42.94	**98.60**
	MEMIT	83.57	79.06	70.52	78.30	77.43	69.44
	F-Learning$_{\text{LoRA}-\text{FT}}$	82.43	77.38	**71.04**	75.17	70.12	69.78
	F-Learning$_{\text{FT}}$	**84.65**	**81.51**	70.92	**83.06**	**79.50**	**70.09**
COUNTERFACT	Original model	18.47	16.95	/	21.61	17.88	/
	LoRA	30.56	23.24	40.08	27.54	21.21	39.75
	F-Learning$_{\text{LoRA}}$	**31.17**	**23.63**	**40.42**	**29.47**	**22.89**	**44.91**
	FT-c	29.23	19.32	19.70	26.97	17.90	20.09
	Full-FT	65.99	44.08	28.34	32.13	31.95	32.51
	ROME	18.41	17.20	**93.60**	21.83	19.08	**92.27**
	MEMIT	61.94	37.45	21.90	**56.94**	31.48	25.70
	F-Learning$_{\text{LoRA}-\text{FT}}$	**78.73**	**51.67**	**29.49**	32.43	26.89	**37.14**
	F-Learning$_{\text{FT}}$	69.53	45.56	28.41	56.39	**39.75**	31.87

Table 1: Results on three metrics of the two datasets based on LLAMA2-7B and LLAMA-7B.

{**"instruction"**: "What city did Marl Young live when he died?", **"input"**: "", **"output"**: "New Orleans" }

5.2 Baselines

To evaluate the effectiveness of the proposed F-Learning method, we conducted experiments on fine-tuning methods and locate-based methods. For fine-tuning methods, we first compare with the full fine-tuning (**Full-FT**) and **LoRA** (Hu et al., 2021), respectively. LoRA (Low-Rank Adaptation) is a technique for fine-tuning large pre-trained language models by introducing small, trainable matrices into each layer of the model's architecture, allowing for efficient adaptation while keeping the majority of the model's parameters frozen. Then we experiment with a fine-tuning approach (**FT-c**) (Zhu et al., 2020) that leverages L_∞ constraint to retain old irrelevant knowledge. For locate-based methods, we first experiment with **ROME** (Meng et al., 2022a), a method updating specific factual associations with causal intervention. Finally, we compare with the **MEMIT** (Meng et al., 2022b) which is a effective method to directly update large-scale memories.

5.3 Completion Details

We use LLAMA2-7B and LLAMA-7B as the base models for our experiments. We are mainly evaluating the ability to update old knowledge to new knowledge, thus we trained the base model on the old knowledge for 3 epochs by full fine-tuning as the **original model** in our experiment (as the same as **original model** in other experiments). Original model has fully learned old knowledge, which makes the forgetting operation reasonable and necessary. To ensure the uniqueness of the model output, we used the greedy decoding strategy during testing. On the hardware side, a total of 4 × A100-80G GPUs were used for the experiments. More details about the experimental settings can be found in the appendix A.2.

5.4 Experimental Results

The experimental results are presented in Table 1, indicating a notable enhancement in learning following the initial forgetting process, irrespective of whether full fine-tuning or LoRA is employed. We obtain the promising results as our F-Learning method outperforming other baselines in most cases. Specifically, compared with other editors, FT-c has only small improvements over the original model. This could be attributed to its norm regularization, which makes FT-c tend to retain a