

Online Adaptation of Language Models with a Memory of Amortized Contexts

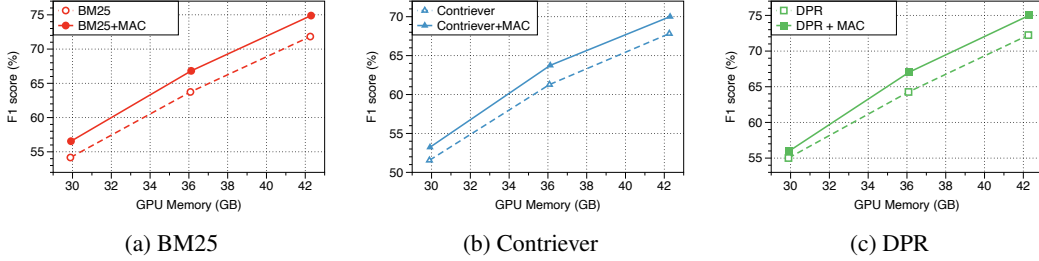


Figure 1: F1 score (%) of various Retrieval-Augmented Generation (RAG) methods (a) BM25, (b) Contriever, and (c) DPR. Here, we report the performance of MAC combined with RAGs. Each point indicates the use of the top-1, top-3, and top-5 retrieved documents, where the x-axis indicates the GPU memory utilization (GB).

Table 1: Detailed values of Exact Match (EM) and F1 scores (%) for various Retrieval-Augmented Generation (RAG) methods, including BM25, Contriever, and DPR. The performance of Memory-Augmented Context (MAC) combined with these RAG methods is also reported. Bold values indicate the best result within each group.

	Top-1		Top-3		Top-5	
	EM	F1	EM	F1	EM	F1
BM25	48.53	54.17	56.18	63.74	64.74	71.83
BM25+MAC	52.81	56.55	60.22	66.82	68.85	74.89
Contriever	44.78	51.55	52.56	61.28	60.10	67.83
Contriever+MAC	47.99	53.23	53.92	63.75	61.28	70.01
DPR	48.98	55.01	57.02	64.27	65.07	72.24
DPR+MAC	49.57	55.98	60.19	67.05	68.52	75.00