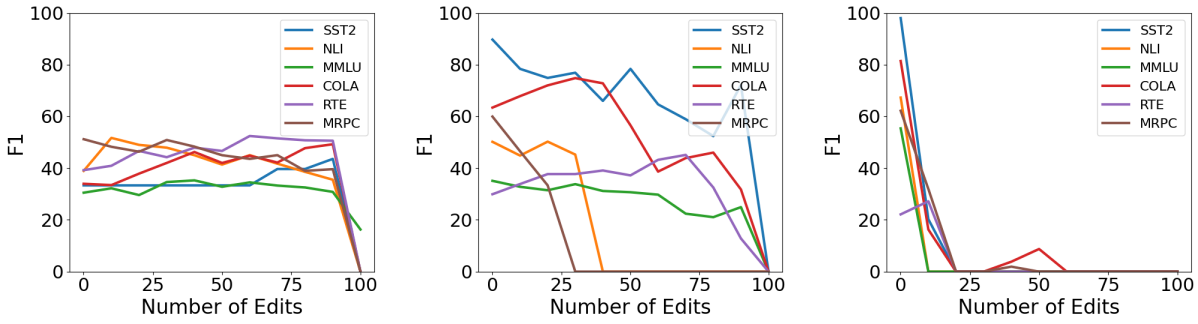
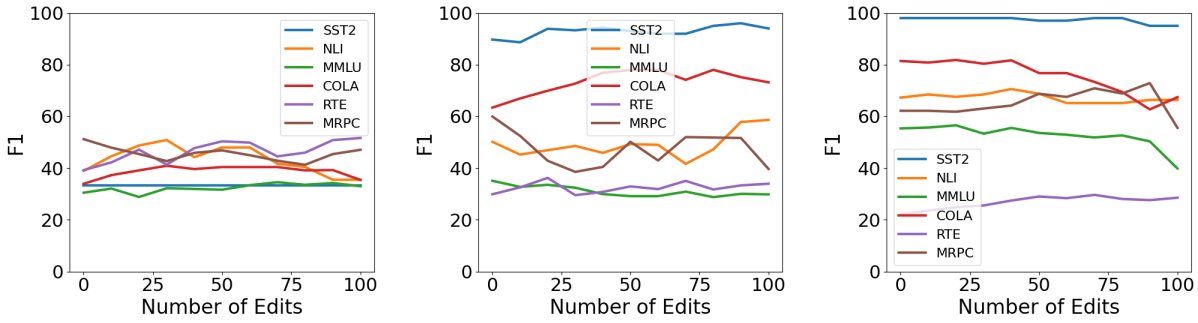
Table 2. Editing performance for PRUNE

METHOD	MODEL	EDIT SCORE	PARAPHRASE SCORE	NEIGHBORHOOD SCORE	OVERALL SCORE	GENERATION ENTROPY
PRUNE	GPT2-XL	61.05	58.05	50.0	55.963	579.69
ENCORE	GPT2-XL	93.21	78.04	59.95	74.58	524.34
PRUNE	LLAMA2-7B	70.8	62.11	51.86	60.597	280.83
ENCORE	LLAMA2-7B	92.57	82.64	60.43	76.043	560.16
PRUNE	LLAMA3-8B	49.38	49.63	51.09	50.022	340.22
ENCORE	LLAMA3-8B	88.77	78.19	60.07	73.707	523.61



(a) Downstream Performance for GPT2-XL using PRUNE (b) Downstream Performance for Llama2-7B using PRUNE (c) Downstream Performance for Llama3-8B using PRUNE

Figure 2. Downstream Performance using PRUNE for different model with CounterFact dataset



(a) Downstream Performance for GPT2-XL using ENCORE (b) Downstream Performance for Llama2-7B using ENCORE (c) Downstream Performance for Llama3-8B using ENCORE

Figure 3. Downstream Performance using ENCORE for different model with CounterFact dataset