METHOD	MODEL	EDIT SCORE	PARAPHRASE SCORE	NEIGHBORHOOD SCORE	OVERALL SCORE	GENERATION ENTROPY
EMMET BASELINE	LLAMA3-8B	85.03	75.06	49.08	65.995	567.34
EMMET + NORM-CONST	LLAMA3-8B	85.52	83.23	50.95	69.228	433.47
EMMET + MPES	LLAMA3-8B	92.96	86.12	55.74	74.424	564.82
EMMET + ENCORE	LLAMA3-8B	92.27	89.46	54.1	74.074	576.89

(a) Downstream Performance for Llama3-8B using EMMET

(b) Downstream Performance for Llama3-8B using EMMET with Norm Constraint

(c) Downstream Performance for Llama3-8B using EMMET with MPES

(d) Downstream Performance for Llama3-8B using EMMET with ENCORE
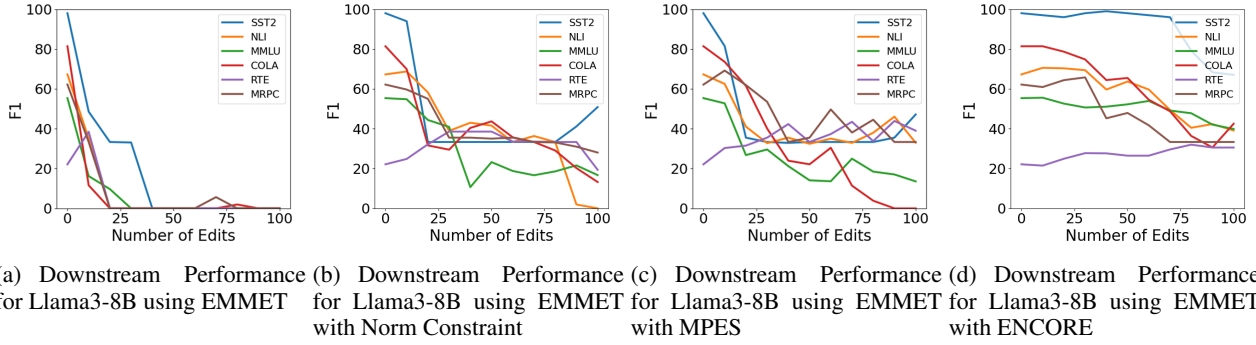
Figure 1. Downstream Performance for Llama3-8B using EMMET, MPES, Norm Constraint and ENCORE with CounterFact dataset

Table 2. Editing performance for EMMET with GPT2-XL with ENCORE.

METHOD	MODEL	EDIT SCORE	PARAPHRASE SCORE	NEIGHBORHOOD SCORE	OVERALL SCORE	GENERATION ENTROPY
EMMET BASELINE	GPT2-XL	79.27	67.45	52.39	64.477	570.24
EMMET + NORM-CONST	GPT2-XL	89.96	71.27	61.23	72.326	592.81
EMMET + MPES	GPT2-XL	95.08	79.47	56.33	73.433	555.39
EMMET + ENCORE	GPT2-XL	87.99	69.03	62.27	71.581	591.17

(a) Downstream Performance for GPT2-XL using EMMET (b) Downstream Performance for GPT2-XL using EMMET with Norm Constraint (c) Downstream Performance for GPT2-XL using EMMET with MPES (d) Downstream Performance for GPT2-XL using EMMET with ENCORE
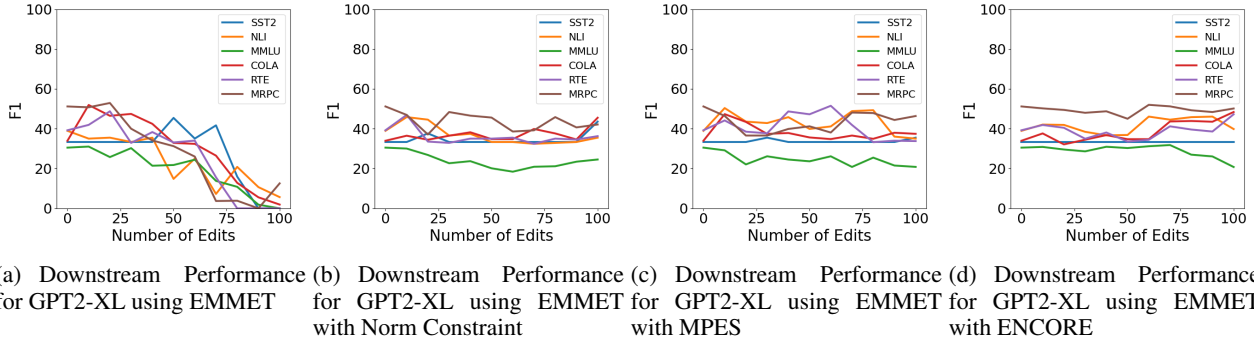
Figure 2. Downstream Performance for GPT2-XL using EMMET, MPES, Norm Constraint and ENCORE with CounterFact dataset