NYU

AdaptEase

Cross-Domain Fine-Tuning for Efficient LMs

Arun Purohit, Aryan Tomar, Moulik Shah, Nimit Kanani, Prashant Shihora

Task 1

Comparison of various parameter efficient fine tuning techniques on T5-large for cross domain sentiment analysis

Comparison against fully Fine tuned Flan-T5-large

Task 2

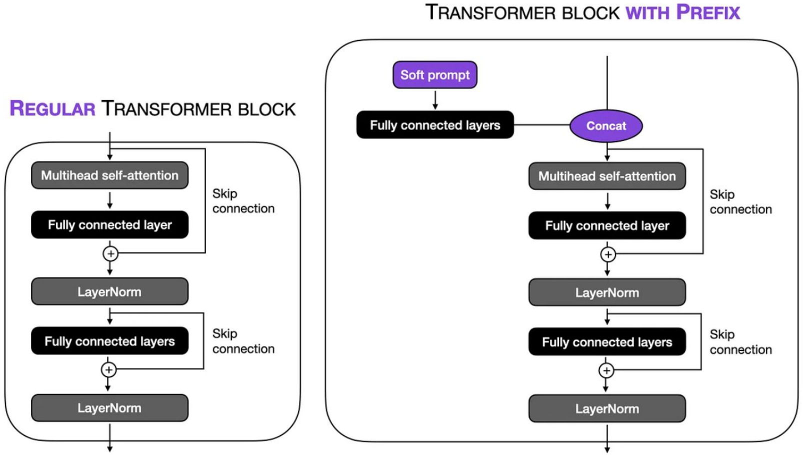
Combining 2 PEFT techniques using a model of even lower parameters for the same cross domain task

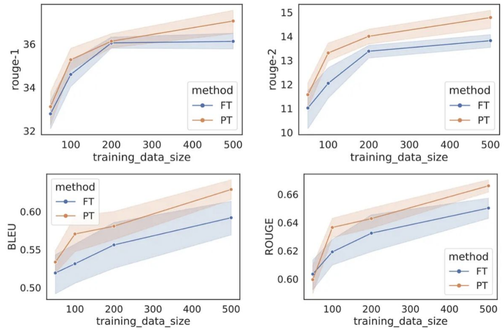
Hypothesis:

Combination of techniques enable lower parametric methods to match the performance of bigger LM

NYU

Soft Prompts: Prefix Tuning





On Low Data Settings on Table-2-Text and Summarization

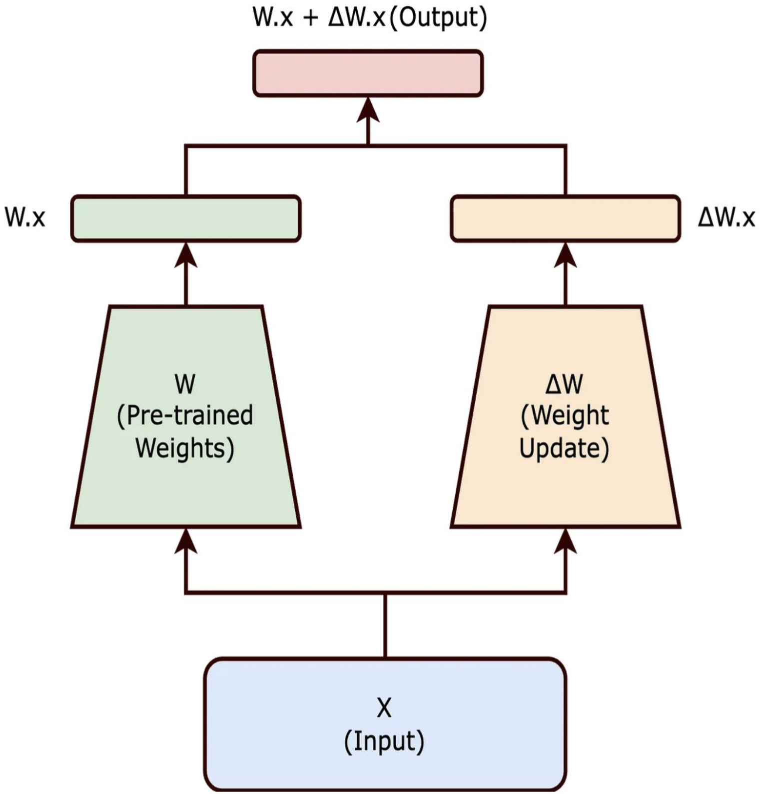
Our Experiment: Cross domain sentiment analysis

Dataset	Prefix Tuned T5-Large	Flan-T5 Zero-shot
SST-2 (Validation)	92.66	**94.72**
IMDB Movie Sentiment	**96.4**	92.4
Financial Phrasebank	**92.91**	89.19
Twitter Sentiment	91.66	**93.2**

Table 2: Model Performance Comparison

NYU

Adapter Based (LoRA)



Dataset	T5-Large LoRA	Prefix-Tuned T5-Large
SST-2 (Validation)	**94.6**	92.66
IMDB Movie Sentiment	95.2	**96.4**
Financial Phrasebank	84.77	**92.91**
Twitter Sentiment	88.57	**93.2**

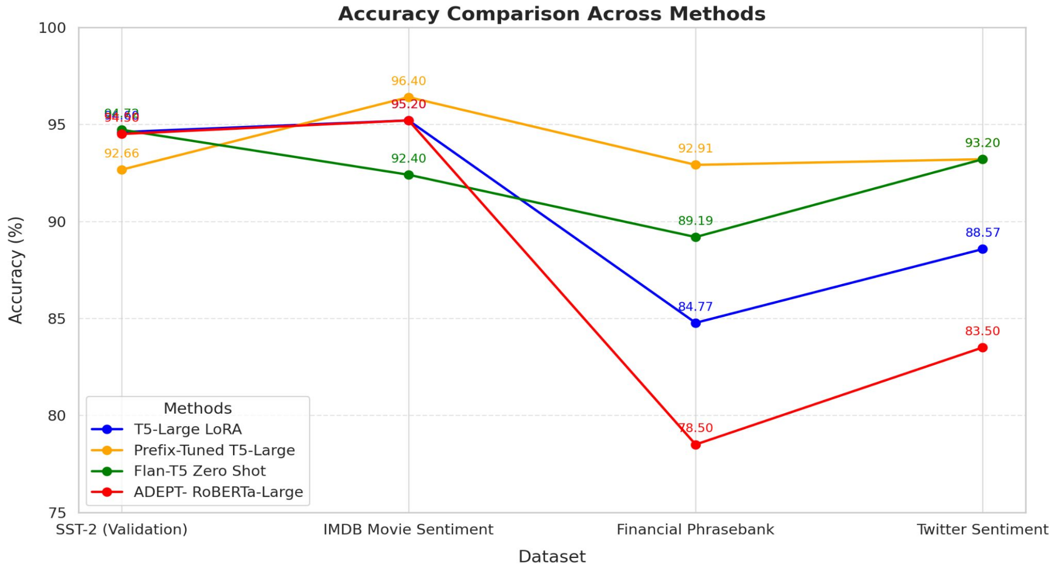
LoRA training parameters:

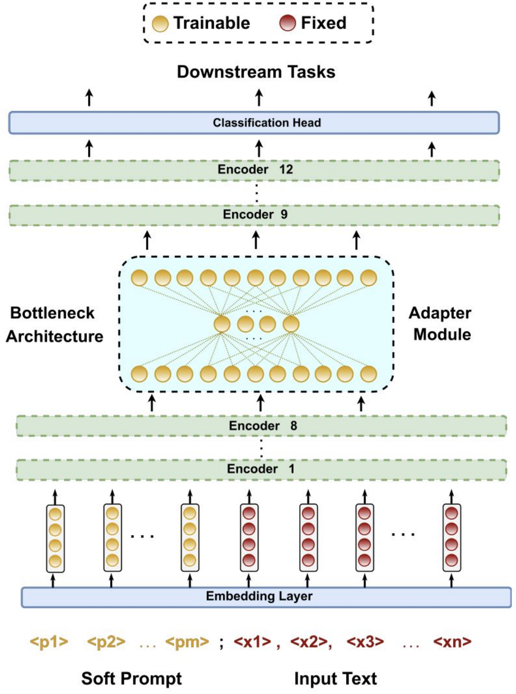
Rank	8
Alpha	16
Epochs	1
Learning Rate	2e-4
Batch Size	64

ADEPT: Adapter+Soft Prompts

Claim: Combination of Adapter and Soft Prompt techniques (from ADEPT) can be equally effective on a lower parametric LM than individual technique on a higher parametric LM

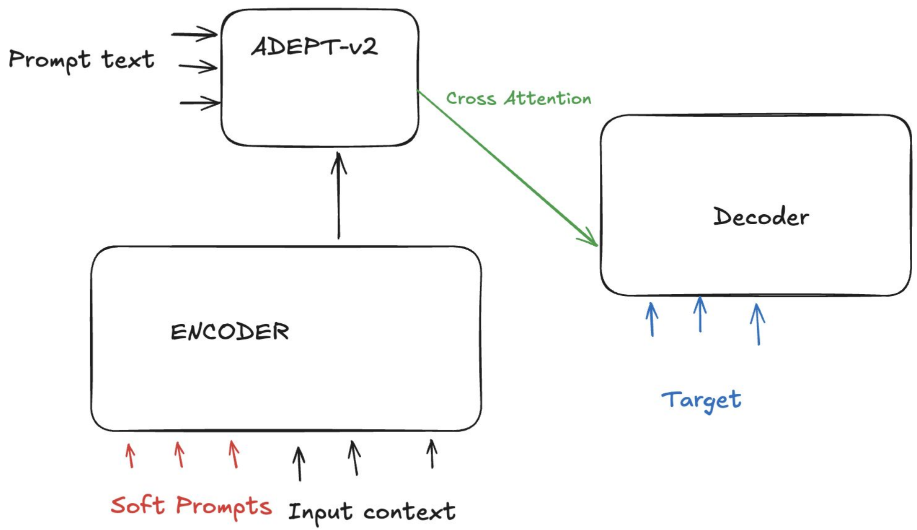
Lower Parametric LM: **RoBERTa-Large** (355 M)
Higher Parametric LM: **T5-large** (770 M)





Future Work

ADEPT in Encoder Decoder: From Cross domain to Cross tasks and many more



Hypothesis:Prompt text helps differentiate between the tasks, Decoder only models perform well across tasks (GPT, LLaMA). However the noise in the input context can cause model hallucinations, hence we want to incorporate the behavior of decoder only models restricting noise in input text allowing creativity via prompt