CS6741 Johnna Liu, jl4528

Tried: bs16/32/128, lr 3e-4/3e-5, epoch 10/20/30/50

Task 1 Finetune

Best performance (in test): bs_16_lr_3e-4_wd_0_epoch_10

	F1 Score	Error Msg Rate
test	0.502314812	0.386574074

Further refinement via:

1. Partly freeze model parameters: esp. lower layers, can reduce overfitting risks and decrease training time; can also apply different learning rates to different layers
2. RAG: generating SQL queries in diverse formats such as applying synonym replacements or syntactically different queries (e.g. GPT3.5)
3. Customed tokenizer: for the identified important components, or training a tokenizer on SQL queries dataset
4. Hyperparameter tuning: try different learning rate, batch size *(although I only found bs16 works, bs32/64/128 all failed)*, warmup steps etc.
5. Error analysis: identify common errors and to refine the structure accordingly

Task 2 Scratch

Best performance (in test): bs_16_lr_3e-4_wd_0_epoch_10

	F1 Score	Error Msg Rate
test	0.546296294	0.194444444

Further refinement via:

1. Incremental training: start from smaller dataset and fewer epochs
2. Learning rate scheduler: e.g. cosine annealing, warmup
3. Initial weight design: e.g. Xavier or He initialization
4. Regularization techniques: e.g. Dropout or L2
5. Model architecture adjustments: try different number of neurons or introducing residual connections

Task 3 Prompt Engineering (Partially Finished)

Till now best performance should be: Few shots + “Let’s think step by step” ☺ + refine some prompts via GPT.