

LowResourceTableQA

Datasets

- Download synthetic BanglaTabQA data (training+validation) at [BanglaTabQA dataset](#)
- Download manually annotated BanglaTabQA [test](#) data
- Download [HindiTabQA dataset](#)
- Download manually annotated HindiTabQA [test](#) data

Model Checkpoints

- **BanglaTabQA Models**
 - Download [BnTQA-mBart](#)
 - Download [BnTQA-M2M](#)
 - Download [BnTQA-llama](#)
- **HindiTabQA Models**
 - Download [HiTQA-mBart](#)
 - Download [HiTQA-M2M](#)
 - Download [HiTQA-llama](#)
 - Download [HiTQA-BnTQA](#)

Bengali SQL query creation

```
python data_generation/extract_wikitable.py --table_language "bn" --
data_save_path "data/bengali_tables.jsonl" --max_table_cells 500
python data_generation/create_sql_samples.py --table_language "bn" --
data_save_path "data/bengali_tables.jsonl" --max_table_cells 500
python data_generation/process_code_mixed_sql.py --input_file
"data/bengali_sql/non_numeric_code_mixed.jsonl" \
--output_file
"data/bengali_sql/non_numeric_full_indic.jsonl" \
--table_language "bn" --
sql_language "bengali"
```

Data Generation Training Process: SQL2NQ

```
python train_sql2nq.py --pretrained_model_name "facebook/mbart-large-50" \
--learning_rate 1e-4 --train_batch_size 4 --eval_batch_size
4 --gradient_accumulation_steps 64 --num_train_epochs 60 \
--use_multiprocessing False --num_workers 2 --
decoder_max_length 1024 --language "bn" \
--output_dir "models/sql2nq" --seed 45 \
--save_total_limit 1
```

Training Process: Low-Resource Table QA

Arguments for Bengali TableQA training:

```
python tableqa/train.py --pretrained_language_model "facebook/mbart-large-50" --learning_rate 1e-4 \
    --train_batch_size 2 --eval_batch_size 2 --
gradient_accumulation_steps 64 --num_train_epochs 8 \
    --use_multiprocessing False --num_workers 2 --
decoder_max_length 1024 \
    --seed 42 --decoder_max_length 1024 --language "bn" \
    --output_dir "experiments/banglaTabQA_mbart"
```

Arguments for Hindi TableQA training:

```
python tableqa/train.py --pretrained_language_model "facebook/mbart-large-50" --learning_rate 1e-4 \
    --train_batch_size 2 --eval_batch_size 2 --
gradient_accumulation_steps 64 --num_train_epochs 8 \
    --use_multiprocessing False --num_workers 2 --
decoder_max_length 1024 \
    --seed 42 --decoder_max_length 1024 --language "hi" \
    --output_dir "experiments/hindiTabQA_mbart"
```

Arguments for Bengali Table QA evaluation:

```
python tableqa/evaluate_tableqa.py --pretrained_model_name
"experiments/banglaTabQA_mbart/latest-checkpoint" \
    --batch_size 2 --generation_max_length 1024 \
    --validation_dataset_path "data/mbart-50_tokenized/mbart-
50_validation.hf" \
    --predictions_save_path "experiments/predictions/mbart-
50_validation.jsonl"
```