

4_0_Model_Training_Results,_Log_File_Parsing_and_High_Level_Analy

April 13, 2025

```
[4]: !sudo apt-get update
      ! sudo apt-get install tree
```

```
Hit:1 https://cloud.r-project.org/bin/linux/ubuntu jammy-cran40/ InRelease
Hit:2 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64
InRelease
Hit:3 http://security.ubuntu.com/ubuntu jammy-security InRelease
Hit:4 https://r2u.stat.illinois.edu/ubuntu jammy InRelease
Hit:5 http://archive.ubuntu.com/ubuntu jammy InRelease
Hit:6 http://archive.ubuntu.com/ubuntu jammy-updates InRelease
Hit:7 http://archive.ubuntu.com/ubuntu jammy-backports InRelease
Hit:8 https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu jammy InRelease
Hit:9 https://ppa.launchpadcontent.net/graphics-drivers/ppa/ubuntu jammy
InRelease
Hit:10 https://ppa.launchpadcontent.net/ubuntugis/ppa/ubuntu jammy InRelease
Reading package lists... Done
W: Skipping acquire of configured file 'main/source/Sources' as repository
'https://r2u.stat.illinois.edu/ubuntu jammy InRelease' does not seem to provide
it (sources.list entry misspelt?)
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
tree is already the newest version (2.0.2-1).
0 upgraded, 0 newly installed, 0 to remove and 31 not upgraded.
```

```
[5]: import re
      import os
      import pandas as pd
      import numpy as np
      import matplotlib.pyplot as plt
      import seaborn as sns
      import json
      import datetime
      from datetime import datetime
      import zoneinfo
```

```
[6]: from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

```
[7]: dir_root = '/content/drive/MyDrive/266-final/'
# dir_data = '/content/drive/MyDrive/266-final/data/'
# dir_data = '/content/drive/MyDrive/266-final/data/se21-t1-comp-lex-master/'
dir_data = '/content/drive/MyDrive/266-final/data/266-comp-lex-master'
dir_models = '/content/drive/MyDrive/266-final/models/'
dir_results = '/content/drive/MyDrive/266-final/results/'
log_filename = "experiment_runs.txt"
log_filepath = os.path.join(dir_results, log_filename)
```

```
[8]: !tree -L 1 /content/drive/MyDrive/266-final/
```

```
/content/drive/MyDrive/266-final/
  data
  models
  notebook-scripts
  paper
  results
  slides
```

6 directories, 0 files

```
[9]: def parse_log_file(file_path):
    """
    Returns a Pandas DataFrame of parsed JSON records.
    """
    records = []
    buffer = []
    brace_depth = 0

    with open(file_path, 'r', encoding='utf-8') as f:
        for line in f:
            brace_diff = line.count('{') - line.count('}')
            brace_depth += brace_diff

            buffer.append(line)

            if brace_depth == 0 and buffer:
                record_str = "".join(buffer).strip()

                if record_str:
                    try:
                        record = json.loads(record_str)
                        records.append(record)
```

```

        except json.JSONDecodeError:
            pass

        buffer = []

    df = pd.json_normalize(records)

    return df

```

```

[10]: log_df = parse_log_file(log_filepath)
      display(log_df)

```

| | timestamp | experiment_meta.model_name \ |
|-----|----------------------------------|------------------------------|
| 0 | 2025-04-09T10:57:21.454889-07:00 | bert-base-cased |
| 1 | 2025-04-09T10:58:27.004579-07:00 | bert-base-cased |
| 2 | 2025-04-09T10:59:19.669997-07:00 | bert-base-cased |
| 3 | 2025-04-09T11:00:14.558023-07:00 | bert-base-cased |
| 4 | 2025-04-09T11:01:07.013948-07:00 | bert-base-cased |
| .. | ... | ... |
| 243 | 2025-04-12T08:02:35.358728-07:00 | answerdotai/ModernBERT-base |
| 244 | 2025-04-12T09:27:28.661745-07:00 | answerdotai/ModernBERT-base |
| 245 | 2025-04-12T09:33:04.239733-07:00 | answerdotai/ModernBERT-base |
| 246 | 2025-04-12T10:00:57.400798-07:00 | answerdotai/ModernBERT-base |
| 247 | 2025-04-12T10:56:47.131248-07:00 | answerdotai/ModernBERT-base |

| | experiment_meta.learning_rate | experiment_meta.epochs \ |
|-----|-------------------------------|--------------------------|
| 0 | 0.000005 | 1 |
| 1 | 0.000005 | 1 |
| 2 | 0.000005 | 1 |
| 3 | 0.000005 | 1 |
| 4 | 0.000005 | 1 |
| .. | ... | ... |
| 243 | 0.000500 | 10 |
| 244 | 0.000500 | 25 |
| 245 | 0.000500 | 5 |
| 246 | 0.000500 | 10 |
| 247 | 0.075000 | 7 |

| | experiment_meta.batch_size | experiment_meta.weight_decay \ |
|-----|----------------------------|--------------------------------|
| 0 | 128 | 0.5000 |
| 1 | 128 | 0.5000 |
| 2 | 128 | 0.5000 |
| 3 | 128 | 0.5000 |
| 4 | 128 | 0.5000 |
| .. | ... | ... |
| 243 | 32 | 0.0000 |
| 244 | 16 | 0.0500 |
| 245 | 16 | 0.0100 |

| | | |
|-----|---|--------|
| 246 | 4 | 0.0025 |
| 247 | 2 | 0.0025 |

| | experiment_meta.x_task | experiment_meta.x_col | experiment_meta.y_col | \ |
|-----|------------------------|-----------------------|-----------------------|---|
| 0 | single | pos_sequence | binary_complexity | |
| 1 | multi | pos_sequence | binary_complexity | |
| 2 | single | morph_sequence | binary_complexity | |
| 3 | multi | morph_sequence | binary_complexity | |
| 4 | single | sentence | binary_complexity | |
| .. | ... | ... | ... | |
| 243 | single | snc_pos_seq | binary_complexity | |
| 244 | single | snc_pos_seq | binary_complexity | |
| 245 | single | snc_pos_seq | binary_complexity | |
| 246 | single | snc_pos_seq | binary_complexity | |
| 247 | single | snc_pos_seq | binary_complexity | |

| | experiment_meta.layers_to_unfreeze | ... | \ |
|-----|---|-----|---|
| 0 | [bert.encoder.layer.0., bert.encoder.layer.11... | ... | |
| 1 | [bert.encoder.layer.0., bert.encoder.layer.11... | ... | |
| 2 | [bert.encoder.layer.0., bert.encoder.layer.11... | ... | |
| 3 | [bert.encoder.layer.0., bert.encoder.layer.11... | ... | |
| 4 | [bert.encoder.layer.0., bert.encoder.layer.11... | ... | |
| .. | ... | ... | |
| 243 | [model.layers.21.attn_norm.weight, model.layer... | ... | |
| 244 | [model.layers.21.attn_norm.weight, model.layer... | ... | |
| 245 | [model.layers.16.attn_norm.weight, model.layer... | ... | |
| 246 | [model.layers.16.attn_norm.weight, model.layer... | ... | |
| 247 | [model.layers.16.attn_norm.weight, model.layer... | ... | |

| | model_details.config_attributes.summary_type | \ |
|-----|--|---|
| 0 | NaN | |
| 1 | NaN | |
| 2 | NaN | |
| 3 | NaN | |
| 4 | NaN | |
| .. | ... | |
| 243 | NaN | |
| 244 | NaN | |
| 245 | NaN | |
| 246 | NaN | |
| 247 | NaN | |

| | model_details.config_attributes.summary_use_proj | \ |
|---|--|---|
| 0 | NaN | |
| 1 | NaN | |
| 2 | NaN | |
| 3 | NaN | |
| 4 | NaN | |

| | |
|-----|-----|
| .. | ... |
| 243 | NaN |
| 244 | NaN |
| 245 | NaN |
| 246 | NaN |
| 247 | NaN |

| | |
|-----|--|
| | model_details.config_attributes.summary_activation \ |
| 0 | NaN |
| 1 | NaN |
| 2 | NaN |
| 3 | NaN |
| 4 | NaN |
| .. | ... |
| 243 | NaN |
| 244 | NaN |
| 245 | NaN |
| 246 | NaN |
| 247 | NaN |

| | |
|-----|--|
| | model_details.config_attributes.summary_last_dropout \ |
| 0 | NaN |
| 1 | NaN |
| 2 | NaN |
| 3 | NaN |
| 4 | NaN |
| .. | ... |
| 243 | NaN |
| 244 | NaN |
| 245 | NaN |
| 246 | NaN |
| 247 | NaN |

| | |
|-----|---|
| | model_details.config_attributes.start_n_top \ |
| 0 | NaN |
| 1 | NaN |
| 2 | NaN |
| 3 | NaN |
| 4 | NaN |
| .. | ... |
| 243 | NaN |
| 244 | NaN |
| 245 | NaN |
| 246 | NaN |
| 247 | NaN |

| | |
|---|---|
| | model_details.config_attributes.end_n_top \ |
| 0 | NaN |

| | |
|-----|-----|
| 1 | NaN |
| 2 | NaN |
| 3 | NaN |
| 4 | NaN |
| .. | ... |
| 243 | NaN |
| 244 | NaN |
| 245 | NaN |
| 246 | NaN |
| 247 | NaN |

| | |
|-----|---|
| | model_details.config_attributes.use_mems_eval \ |
| 0 | NaN |
| 1 | NaN |
| 2 | NaN |
| 3 | NaN |
| 4 | NaN |
| .. | ... |
| 243 | NaN |
| 244 | NaN |
| 245 | NaN |
| 246 | NaN |
| 247 | NaN |

| | |
|-----|--|
| | model_details.config_attributes.use_mems_train \ |
| 0 | NaN |
| 1 | NaN |
| 2 | NaN |
| 3 | NaN |
| 4 | NaN |
| .. | ... |
| 243 | NaN |
| 244 | NaN |
| 245 | NaN |
| 246 | NaN |
| 247 | NaN |

| | |
|-------------|---|
| | model_details.config_attributes.task_specific_params.text-generation. |
| do_sample \ | |
| 0 | NaN |
| 1 | NaN |
| 2 | NaN |
| 3 | NaN |
| 4 | NaN |
| .. | ... |
| 243 | NaN |
| 244 | NaN |

| | |
|-----|-----|
| 245 | NaN |
| 246 | NaN |
| 247 | NaN |

```

    model_details.config_attributes.task_specific_params.text-generation.
    ↪max_length
0      NaN
1      NaN
2      NaN
3      NaN
4      NaN
..     ...
243    NaN
244    NaN
245    NaN
246    NaN
247    NaN

```

[248 rows x 180 columns]

```
[11]: log_df.shape
```

```
[11]: (248, 180)
```

```
[12]: !tree /content/drive/MyDrive/266-final/results/
```

```

/content/drive/MyDrive/266-final/results/
├── data_engineering_qa
│   ├── sentence_no_contraction_span_analysis.csv
│   ├── sentence_span_analysis (1).csv
│   ├── sentence_span_analysis.csv
│   ├── sentence_span_analysis_no_contractions (1).csv
│   └── sentence_span_analysis_no_contractions.csv
├── experiment_results_2025-04-11T09:40:45.673867-07:00.xlsx
├── experiment_results_2025-04-11T15:09:24.583184-07:00.xlsx
└── experiment-results-archive
    ├── archive2_experiment_runs.txt
    ├── archive3_experiment_runs.txt
    ├── archive_experiment_runs.txt
    ├── Copy of experiment_runs.txt
    ├── experiment_results_20250410_034821.xlsx
    ├── experiment_results_20250410_164111.xlsx
    ├── experiment_results_20250410_165208.xlsx
    ├── experiment_results_20250410_223015.xlsx
    ├── experiment_results_20250410_223135.xlsx
    ├── experiment_results_20250410_223248.xlsx
    ├── experiment_results_20250410_224758.xlsx
    └── experiment_results_20250410_232025.xlsx

```

```

experiment_results_20250411_000318.xlsx
experiment_results_20250411_003914.xlsx
experiment_results_20250411_010722.xlsx
experiment_results_20250411_051357.xlsx
experiment_results_20250411_121251.xlsx
experiment_results_20250411_123253.xlsx
experiment_results_20250411_124625.xlsx
experiment_results_20250411_125911.xlsx
experiment_results_20250411_162910.xlsx
experiment_runs.txt
sentence_span_analysis.csv
sentence_span_analysis_no_contractions.csv

```

2 directories, 31 files

```

[13]: with open(log_filepath, 'r', encoding='utf-8') as f:
        content = f.read()
        print(content)

```

```

{"timestamp": "2025-04-09T10:57:21.454889-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.0.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 14767874, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":

```



```

"sdpa", "attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.692979633808136, "train_eval_accuracy": 0.5125293657008614,
"train_eval_precision": 0.5052435723951285, "train_eval_recall":
0.7866736897550698, "train_eval_f1": 0.6153053867545576, "train_eval_runtime":
13.1963, "train_eval_samples_per_second": 580.616,
"train_eval_steps_per_second": 4.547, "train_epoch": 1.0, "val_eval_loss":
0.692278265953064, "val_eval_accuracy": 0.5320665083135392,
"val_eval_precision": 0.49240121580547114, "val_eval_recall": 0.84375,
"val_eval_f1": 0.6218809980806143, "val_eval_runtime": 2.1541,
"val_eval_samples_per_second": 195.437, "val_eval_steps_per_second": 1.857,
"val_epoch": 1.0, "test_eval_loss": 0.6909976601600647, "test_eval_accuracy":
0.520174482006543, "test_eval_precision": 0.5007256894049347,
"test_eval_recall": 0.782312925170068, "test_eval_f1": 0.6106194690265486,
"test_eval_runtime": 2.6416, "test_eval_samples_per_second": 347.136,
"test_eval_steps_per_second": 3.028, "test_epoch": 1.0}}
{"timestamp": "2025-04-09T10:58:27.004579-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.0.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 14767874, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,

```

```

"task_specific_params": null, "problem_type": "single_label_classification",
"name_or_path": "bert-base-cased", "commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "attn_implementation_internal":
"sdpa", "attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6902915239334106, "train_eval_accuracy": 0.5390237535891412,
"train_eval_precision": 0.5463773188659433, "train_eval_recall":
0.4111140373979457, "train_eval_f1": 0.46919146378118426, "train_eval_runtime":
12.9643, "train_eval_samples_per_second": 591.006,
"train_eval_steps_per_second": 4.628, "train_epoch": 1.0, "val_eval_loss":
0.6824801564216614, "val_eval_accuracy": 0.5795724465558195,
"val_eval_precision": 0.5483870967741935, "val_eval_recall": 0.4427083333333333,
"val_eval_f1": 0.4899135446685879, "val_eval_runtime": 3.176,
"val_eval_samples_per_second": 132.559, "val_eval_steps_per_second": 1.259,
"val_epoch": 1.0, "test_eval_loss": 0.6860960125923157, "test_eval_accuracy":
0.5561613958560524, "test_eval_precision": 0.555921052631579,
"test_eval_recall": 0.3832199546485261, "test_eval_f1": 0.45369127516778524,
"test_eval_runtime": 4.7886, "test_eval_samples_per_second": 191.498,
"test_eval_steps_per_second": 1.671, "test_epoch": 1.0}}
{"timestamp": "2025-04-09T10:59:19.669997-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.0.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 14767874, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":

```

```

"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7113794088363647, "train_eval_accuracy": 0.49569303054032887,
"train_eval_precision": 0.4956260608434521, "train_eval_recall":
0.9997366341848828, "train_eval_f1": 0.6627094972067039, "train_eval_runtime":
13.4141, "train_eval_samples_per_second": 571.191,
"train_eval_steps_per_second": 4.473, "train_epoch": 1.0, "val_eval_loss":
0.7271953225135803, "val_eval_accuracy": 0.45605700712589076,
"val_eval_precision": 0.45605700712589076, "val_eval_recall": 1.0,
"val_eval_f1": 0.6264274061990212, "val_eval_runtime": 2.1856,
"val_eval_samples_per_second": 192.625, "val_eval_steps_per_second": 1.83,
"val_epoch": 1.0, "test_eval_loss": 0.7140419483184814, "test_eval_accuracy":
0.48091603053435117, "test_eval_precision": 0.48091603053435117,
"test_eval_recall": 1.0, "test_eval_f1": 0.6494845360824743,
"test_eval_runtime": 2.8863, "test_eval_samples_per_second": 317.704,
"test_eval_steps_per_second": 2.772, "test_epoch": 1.0}}
{"timestamp": "2025-04-09T11:00:14.558023-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.0.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 14767874, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,

```

```

"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7113794088363647, "train_eval_accuracy": 0.49569303054032887,
"train_eval_precision": 0.4956260608434521, "train_eval_recall":
0.9997366341848828, "train_eval_f1": 0.6627094972067039, "train_eval_runtime":
13.479, "train_eval_samples_per_second": 568.441, "train_eval_steps_per_second":
4.451, "train_epoch": 1.0, "val_eval_loss": 0.7271953225135803,
"val_eval_accuracy": 0.45605700712589076, "val_eval_precision":
0.45605700712589076, "val_eval_recall": 1.0, "val_eval_f1": 0.6264274061990212,
"val_eval_runtime": 2.3915, "val_eval_samples_per_second": 176.04,
"val_eval_steps_per_second": 1.673, "val_epoch": 1.0, "test_eval_loss":
0.7140419483184814, "test_eval_accuracy": 0.48091603053435117,
"test_eval_precision": 0.48091603053435117, "test_eval_recall": 1.0,
"test_eval_f1": 0.6494845360824743, "test_eval_runtime": 3.0385,
"test_eval_samples_per_second": 301.796, "test_eval_steps_per_second": 2.633,
"test_epoch": 1.0}}
{"timestamp": "2025-04-09T11:01:07.013948-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col": "sentence",
"y_col": "binary_complexity", "layers_to_unfreeze": ["bert.encoder.layer.0.",
"bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 14767874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,

```

```

"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7113794088363647, "train_eval_accuracy": 0.49569303054032887,
"train_eval_precision": 0.4956260608434521, "train_eval_recall":
0.9997366341848828, "train_eval_f1": 0.6627094972067039, "train_eval_runtime":
12.9378, "train_eval_samples_per_second": 592.219,
"train_eval_steps_per_second": 4.638, "train_epoch": 1.0, "val_eval_loss":
0.7271953225135803, "val_eval_accuracy": 0.45605700712589076,
"val_eval_precision": 0.45605700712589076, "val_eval_recall": 1.0,
"val_eval_f1": 0.6264274061990212, "val_eval_runtime": 3.2155,
"val_eval_samples_per_second": 130.929, "val_eval_steps_per_second": 1.244,
"val_epoch": 1.0, "test_eval_loss": 0.7140419483184814, "test_eval_accuracy":
0.48091603053435117, "test_eval_precision": 0.48091603053435117,
"test_eval_recall": 1.0, "test_eval_f1": 0.6494845360824743,
"test_eval_runtime": 3.9224, "test_eval_samples_per_second": 233.785,
"test_eval_steps_per_second": 2.04, "test_epoch": 1.0}}
{"timestamp": "2025-04-09T11:01:57.744475-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col": "sentence",
"y_col": "binary_complexity", "layers_to_unfreeze": ["bert.encoder.layer.0.",
"bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 14767874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":

```

```

1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7113794088363647, "train_eval_accuracy": 0.49569303054032887,
"train_eval_precision": 0.4956260608434521, "train_eval_recall":
0.9997366341848828, "train_eval_f1": 0.6627094972067039, "train_eval_runtime":
13.1219, "train_eval_samples_per_second": 583.907,
"train_eval_steps_per_second": 4.572, "train_epoch": 1.0, "val_eval_loss":
0.7271953225135803, "val_eval_accuracy": 0.45605700712589076,
"val_eval_precision": 0.45605700712589076, "val_eval_recall": 1.0,
"val_eval_f1": 0.6264274061990212, "val_eval_runtime": 2.068,
"val_eval_samples_per_second": 203.58, "val_eval_steps_per_second": 1.934,
"val_epoch": 1.0, "test_eval_loss": 0.7140419483184814, "test_eval_accuracy":
0.48091603053435117, "test_eval_precision": 0.48091603053435117,
"test_eval_recall": 1.0, "test_eval_f1": 0.6494845360824743,
"test_eval_runtime": 2.4572, "test_eval_samples_per_second": 373.194,
"test_eval_steps_per_second": 3.256, "test_epoch": 1.0}}
{"timestamp": "2025-04-09T11:50:47.507240-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.0.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 14767874, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,

```

```

"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6945602297782898, "train_eval_accuracy": 0.4988253719655442,
"train_eval_precision": 0.4963810806261572, "train_eval_recall":
0.7766657887806163, "train_eval_f1": 0.6056685150955021, "train_eval_runtime":
13.5504, "train_eval_samples_per_second": 565.443,
"train_eval_steps_per_second": 4.428, "train_epoch": 1.0, "val_eval_loss":
0.6970546841621399, "val_eval_accuracy": 0.4513064133016627,
"val_eval_precision": 0.4421364985163205, "val_eval_recall": 0.7760416666666666,
"val_eval_f1": 0.5633270321361059, "val_eval_runtime": 1.8111,
"val_eval_samples_per_second": 232.456, "val_eval_steps_per_second": 2.209,
"val_epoch": 1.0, "test_eval_loss": 0.6963492035865784, "test_eval_accuracy":
0.45910577971646677, "test_eval_precision": 0.46088193456614507,
"test_eval_recall": 0.7346938775510204, "test_eval_f1": 0.5664335664335665,
"test_eval_runtime": 2.9708, "test_eval_samples_per_second": 308.668,
"test_eval_steps_per_second": 2.693, "test_epoch": 1.0}}
{"timestamp": "2025-04-09T11:51:06.095185-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.0.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 14767874, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},

```

```

"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7657436728477478, "train_eval_accuracy": 0.4996704021094265,
"train_eval_precision": 0.4996704021094265, "train_eval_recall": 1.0,
"train_eval_f1": 0.6663736263736264, "train_eval_runtime": 3.5146,
"train_eval_samples_per_second": 431.626, "train_eval_steps_per_second": 3.414,
"train_epoch": 1.0, "val_eval_loss": 0.7580294609069824, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 1.5997,
"val_eval_samples_per_second": 61.888, "val_eval_steps_per_second": 0.625,
"val_epoch": 1.0, "test_eval_loss": 0.7353547811508179, "test_eval_accuracy":
0.5380434782608695, "test_eval_precision": 0.5380434782608695,
"test_eval_recall": 1.0, "test_eval_f1": 0.6996466431095406,
"test_eval_runtime": 2.3083, "test_eval_samples_per_second": 79.712,
"test_eval_steps_per_second": 0.866, "test_epoch": 1.0}}
{"timestamp": "2025-04-09T11:52:02.751112-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.0.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 14767874, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,

```



```

"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6931447386741638, "train_eval_accuracy": 0.5053510832680762,
"train_eval_precision": 0.5008509603695599, "train_eval_recall":
0.5425335791414274, "train_eval_f1": 0.5208596713021492, "train_eval_runtime":
13.9426, "train_eval_samples_per_second": 549.54, "train_eval_steps_per_second":
4.303, "train_epoch": 1.0, "val_eval_loss": 0.691696286201477,
"val_eval_accuracy": 0.5130641330166271, "val_eval_precision":
0.4694835680751174, "val_eval_recall": 0.5208333333333334, "val_eval_f1":
0.49382716049382713, "val_eval_runtime": 2.2476, "val_eval_samples_per_second":
187.314, "val_eval_steps_per_second": 1.78, "val_epoch": 1.0, "test_eval_loss":
0.6932744383811951, "test_eval_accuracy": 0.5038167938931297,
"test_eval_precision": 0.4862204724409449, "test_eval_recall":
0.5600907029478458, "test_eval_f1": 0.5205479452054794, "test_eval_runtime":
4.025, "test_eval_samples_per_second": 227.825, "test_eval_steps_per_second":
1.988, "test_epoch": 1.0}}
{"timestamp": "2025-04-09T11:52:22.318529-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.0.", "bert.encoder.layer.11.", "bert.pooler."],

```

```

"classifier."}], "model_details": {"model_total_params": 108311810,
"model_trainable_params": 14767874, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7560539245605469, "train_eval_accuracy": 0.4996704021094265,
"train_eval_precision": 0.4996704021094265, "train_eval_recall": 1.0,
"train_eval_f1": 0.6663736263736264, "train_eval_runtime": 3.581,
"train_eval_samples_per_second": 423.63, "train_eval_steps_per_second": 3.351,
"train_epoch": 1.0, "val_eval_loss": 0.7434189319610596, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 1.3456,
"val_eval_samples_per_second": 73.573, "val_eval_steps_per_second": 0.743,
"val_epoch": 1.0, "test_eval_loss": 0.7248002886772156, "test_eval_accuracy":
0.5380434782608695, "test_eval_precision": 0.5380434782608695,
"test_eval_recall": 1.0, "test_eval_f1": 0.6996466431095406,
"test_eval_runtime": 3.1696, "test_eval_samples_per_second": 58.051,
"test_eval_steps_per_second": 0.631, "test_epoch": 1.0}}
{"timestamp": "2025-04-09T11:53:23.579514-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col": "sentence",

```

```

"y_col": "binary_complexity", "layers_to_unfreeze": ["bert.encoder.layer.0.",
"bert.encoder.layer.11.", "bert.pooler.", "classifier."], "model_details":
{"model_total_params": 108311810, "model_trainable_params": 14767874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7117305397987366, "train_eval_accuracy": 0.49569303054032887,
"train_eval_precision": 0.4956260608434521, "train_eval_recall":
0.9997366341848828, "train_eval_f1": 0.6627094972067039, "train_eval_runtime":
14.7504, "train_eval_samples_per_second": 519.444,
"train_eval_steps_per_second": 4.068, "train_epoch": 1.0, "val_eval_loss":
0.7273501753807068, "val_eval_accuracy": 0.45605700712589076,
"val_eval_precision": 0.45584725536992843, "val_eval_recall":
0.9947916666666666, "val_eval_f1": 0.6252045826513911, "val_eval_runtime":
1.9973, "val_eval_samples_per_second": 210.78, "val_eval_steps_per_second":
2.003, "val_epoch": 1.0, "test_eval_loss": 0.7145293951034546,
"test_eval_accuracy": 0.48091603053435117, "test_eval_precision":
0.48091603053435117, "test_eval_recall": 1.0, "test_eval_f1":
0.6494845360824743, "test_eval_runtime": 2.4838, "test_eval_samples_per_second":
369.199, "test_eval_steps_per_second": 3.221, "test_epoch": 1.0}}
{"timestamp": "2025-04-09T11:53:42.159853-07:00", "experiment_meta":

```

```

{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col": "sentence",
"y_col": "binary_complexity", "layers_to_unfreeze": ["bert.encoder.layer.0.",
"bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 14767874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7740325331687927, "train_eval_accuracy": 0.4996704021094265,
"train_eval_precision": 0.4996704021094265, "train_eval_recall": 1.0,
"train_eval_f1": 0.6663736263736264, "train_eval_runtime": 3.5289,
"train_eval_samples_per_second": 429.874, "train_eval_steps_per_second": 3.4,
"train_epoch": 1.0, "val_eval_loss": 0.7528352737426758, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 1.482,
"val_eval_samples_per_second": 66.804, "val_eval_steps_per_second": 0.675,
"val_epoch": 1.0, "test_eval_loss": 0.739983856678009, "test_eval_accuracy":
0.5380434782608695, "test_eval_precision": 0.5380434782608695,
"test_eval_recall": 1.0, "test_eval_f1": 0.6996466431095406,
"test_eval_runtime": 2.4214, "test_eval_samples_per_second": 75.988,
"test_eval_steps_per_second": 0.826, "test_epoch": 1.0}}

```

```

{"timestamp": "2025-04-09T16:46:20.316052-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 3,
"batch_size": 16, "weight_decay": 0.1, "x_task": "single", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.0.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 14767874, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7185492515563965, "train_eval_accuracy": 0.49595405899243017,
"train_eval_precision": 0.4957510785723624, "train_eval_recall":
0.998683170924414, "train_eval_f1": 0.6625895509348244, "train_eval_runtime":
26.3099, "train_eval_samples_per_second": 291.221,
"train_eval_steps_per_second": 18.206, "train_epoch": 3.0, "val_eval_loss":
0.736166775226593, "val_eval_accuracy": 0.45605700712589076,
"val_eval_precision": 0.45605700712589076, "val_eval_recall": 1.0,
"val_eval_f1": 0.6264274061990212, "val_eval_runtime": 2.4681,
"val_eval_samples_per_second": 170.576, "val_eval_steps_per_second": 10.94,
"val_epoch": 3.0, "test_eval_loss": 0.7259107232093811, "test_eval_accuracy":
0.4820065430752454, "test_eval_precision": 0.4814410480349345,

```

```

"test_eval_recall": 1.0, "test_eval_f1": 0.6499631540162122,
"test_eval_runtime": 4.063, "test_eval_samples_per_second": 225.698,
"test_eval_steps_per_second": 14.275, "test_epoch": 3.0}}
{"timestamp": "2025-04-09T17:05:44.258150-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 10,
"batch_size": 16, "weight_decay": 0.1, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.0.", "bert.encoder.layer.8.", "bert.encoder.layer.9.",
"bert.encoder.layer.10.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 36031490, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.2997378706932068, "train_eval_accuracy": 0.8829287392325763,
"train_eval_precision": 0.9142857142857143, "train_eval_recall":
0.8427706083750329, "train_eval_f1": 0.8770727696313554, "train_eval_runtime":
26.7301, "train_eval_samples_per_second": 286.644,
"train_eval_steps_per_second": 17.92, "train_epoch": 10.0, "val_eval_loss":
0.8265407085418701, "val_eval_accuracy": 0.6080760095011877,
"val_eval_precision": 0.5754189944134078, "val_eval_recall": 0.5364583333333334,

```

```

"val_eval_f1": 0.555256064690027, "val_eval_runtime": 3.2952,
"val_eval_samples_per_second": 127.763, "val_eval_steps_per_second": 8.194,
"val_epoch": 10.0, "test_eval_loss": 0.8846817016601562, "test_eval_accuracy":
0.5768811341330425, "test_eval_precision": 0.5667506297229219,
"test_eval_recall": 0.5102040816326531, "test_eval_f1": 0.5369928400954654,
"test_eval_runtime": 4.1314, "test_eval_samples_per_second": 221.96,
"test_eval_steps_per_second": 14.039, "test_epoch": 10.0}}
{"timestamp": "2025-04-09T17:17:12.024451-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 10,
"batch_size": 16, "weight_decay": 0.1, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.0.", "bert.encoder.layer.8.", "bert.encoder.layer.9.",
"bert.encoder.layer.10.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 36031490, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.2710617184638977, "train_eval_accuracy": 0.8976768467762987,
"train_eval_precision": 0.8859851396361773, "train_eval_recall":
0.9107189886752699, "train_eval_f1": 0.8981818181818182, "train_eval_runtime":

```

```

26.4086, "train_eval_samples_per_second": 290.132,
"train_eval_steps_per_second": 18.138, "train_epoch": 10.0, "val_eval_loss":
0.8370015621185303, "val_eval_accuracy": 0.6104513064133017,
"val_eval_precision": 0.57, "val_eval_recall": 0.59375, "val_eval_f1":
0.5816326530612245, "val_eval_runtime": 2.5325, "val_eval_samples_per_second":
166.236, "val_eval_steps_per_second": 10.661, "val_epoch": 10.0,
"test_eval_loss": 0.9402342438697815, "test_eval_accuracy": 0.5670665212649946,
"test_eval_precision": 0.5491071428571429, "test_eval_recall":
0.5578231292517006, "test_eval_f1": 0.5534308211473565, "test_eval_runtime":
4.0841, "test_eval_samples_per_second": 224.527, "test_eval_steps_per_second":
14.201, "test_epoch": 10.0}}
{"timestamp": "2025-04-09T17:28:50.056819-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 10,
"batch_size": 16, "weight_decay": 0.1, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.0.", "bert.encoder.layer.8.", "bert.encoder.layer.9.",
"bert.encoder.layer.10.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 36031490, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":

```



```

true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6801885366439819, "train_eval_accuracy": 0.568780997128687,
"train_eval_precision": 0.5597286164283983, "train_eval_recall":
0.6083750329207269, "train_eval_f1": 0.5830388692579506, "train_eval_runtime":
27.9135, "train_eval_samples_per_second": 274.491,
"train_eval_steps_per_second": 17.16, "train_epoch": 10.0, "val_eval_loss":
0.6857385635375977, "val_eval_accuracy": 0.5486935866983373,
"val_eval_precision": 0.5043859649122807, "val_eval_recall": 0.5989583333333334,
"val_eval_f1": 0.5476190476190477, "val_eval_runtime": 2.8147,
"val_eval_samples_per_second": 149.57, "val_eval_steps_per_second": 9.592,
"val_epoch": 10.0, "test_eval_loss": 0.6894793510437012, "test_eval_accuracy":
0.5539803707742639, "test_eval_precision": 0.5333333333333333,
"test_eval_recall": 0.5804988662131519, "test_eval_f1": 0.5559174809989142,
"test_eval_runtime": 4.4064, "test_eval_samples_per_second": 208.108,
"test_eval_steps_per_second": 13.163, "test_epoch": 10.0}}
{"timestamp": "2025-04-09T17:31:29.774764-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 10,
"batch_size": 16, "weight_decay": 0.1, "x_task": "multi", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.0.", "bert.encoder.layer.8.", "bert.encoder.layer.9.",
"bert.encoder.layer.10.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 36031490, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,

```

```

"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6708851456642151, "train_eval_accuracy": 0.5952537903757416,
"train_eval_precision": 0.5821917808219178, "train_eval_recall":
0.6728232189973615, "train_eval_f1": 0.6242350061199511, "train_eval_runtime":
6.469, "train_eval_samples_per_second": 234.505, "train_eval_steps_per_second":
14.686, "train_epoch": 10.0, "val_eval_loss": 0.720668375492096,
"val_eval_accuracy": 0.47474747474747475, "val_eval_precision":
0.49122807017543857, "val_eval_recall": 0.5490196078431373, "val_eval_f1":
0.5185185185185185, "val_eval_runtime": 2.1093, "val_eval_samples_per_second":
46.935, "val_eval_steps_per_second": 3.319, "val_epoch": 10.0, "test_eval_loss":
0.72453373670578, "test_eval_accuracy": 0.5108695652173914,
"test_eval_precision": 0.5473684210526316, "test_eval_recall":
0.5252525252525253, "test_eval_f1": 0.5360824742268041, "test_eval_runtime":
2.7948, "test_eval_samples_per_second": 65.837, "test_eval_steps_per_second":
4.294, "test_epoch": 10.0}}
{"timestamp": "2025-04-09T17:43:04.832450-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 10,
"batch_size": 16, "weight_decay": 0.1, "x_task": "single", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.0.", "bert.encoder.layer.8.", "bert.encoder.layer.9.",
"bert.encoder.layer.10.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 36031490, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":

```

```

"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6766506433486938, "train_eval_accuracy": 0.5754372226572696,
"train_eval_precision": 0.5775370581527937, "train_eval_recall":
0.5335791414274427, "train_eval_f1": 0.5546885694729637, "train_eval_runtime":
26.9737, "train_eval_samples_per_second": 284.055,
"train_eval_steps_per_second": 17.758, "train_epoch": 10.0, "val_eval_loss":
0.6723506450653076, "val_eval_accuracy": 0.5748218527315915,
"val_eval_precision": 0.5320197044334976, "val_eval_recall": 0.5625,
"val_eval_f1": 0.5468354430379747, "val_eval_runtime": 2.444,
"val_eval_samples_per_second": 172.261, "val_eval_steps_per_second": 11.048,
"val_epoch": 10.0, "test_eval_loss": 0.6912660598754883, "test_eval_accuracy":
0.5528898582333697, "test_eval_precision": 0.5354691075514875,
"test_eval_recall": 0.5306122448979592, "test_eval_f1": 0.5330296127562643,
"test_eval_runtime": 4.1153, "test_eval_samples_per_second": 222.826,
"test_eval_steps_per_second": 14.094, "test_epoch": 10.0}}
{"timestamp": "2025-04-09T17:45:38.394081-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 10,
"batch_size": 16, "weight_decay": 0.1, "x_task": "multi", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.0.", "bert.encoder.layer.8.", "bert.encoder.layer.9.",
"bert.encoder.layer.10.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 36031490, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},

```

```

"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdp", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6437759399414062, "train_eval_accuracy": 0.6348055372445617,
"train_eval_precision": 0.6416666666666667, "train_eval_recall":
0.6094986807387863, "train_eval_f1": 0.625169147496617, "train_eval_runtime":
6.3442, "train_eval_samples_per_second": 239.117, "train_eval_steps_per_second":
14.974, "train_epoch": 10.0, "val_eval_loss": 0.6641461253166199,
"val_eval_accuracy": 0.6363636363636364, "val_eval_precision":
0.6744186046511628, "val_eval_recall": 0.5686274509803921, "val_eval_f1":
0.6170212765957447, "val_eval_runtime": 1.3673, "val_eval_samples_per_second":
72.407, "val_eval_steps_per_second": 5.12, "val_epoch": 10.0, "test_eval_loss":
0.726580798625946, "test_eval_accuracy": 0.5217391304347826,
"test_eval_precision": 0.582089552238806, "test_eval_recall":
0.3939393939393939, "test_eval_f1": 0.46987951807228917, "test_eval_runtime":
1.8213, "test_eval_samples_per_second": 101.026, "test_eval_steps_per_second":
6.589, "test_epoch": 10.0}}
{"timestamp": "2025-04-09T19:52:49.131288-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 5,
"batch_size": 16, "weight_decay": 0.1, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.embeddings.", "bert.encoder.layer.0.", "bert.encoder.layer.8.",
"bert.encoder.layer.9.", "bert.encoder.layer.10.", "bert.encoder.layer.11.",
"bert.pooler.", "classifier."]}, "model_details": {"model_total_params":
108311810, "model_trainable_params": 58696706, "pooler_activation_function":
"Tanh", "config_attributes": {"return_dict": true, "output_hidden_states":
false, "output_attentions": false, "torchscript": false, "torch_dtype":
"float32", "use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,

```

```

"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.5153685808181763, "train_eval_accuracy": 0.7512398851474811,
"train_eval_precision": 0.7571389719880337, "train_eval_recall":
0.7332104292862787, "train_eval_f1": 0.744982606368745, "train_eval_runtime":
26.5942, "train_eval_samples_per_second": 288.108,
"train_eval_steps_per_second": 18.011, "train_epoch": 5.0, "val_eval_loss":
0.6726545095443726, "val_eval_accuracy": 0.6318289786223278,
"val_eval_precision": 0.5948717948717949, "val_eval_recall": 0.6041666666666666,
"val_eval_f1": 0.599483204134367, "val_eval_runtime": 2.4042,
"val_eval_samples_per_second": 175.11, "val_eval_steps_per_second": 11.23,
"val_epoch": 5.0, "test_eval_loss": 0.7117065787315369, "test_eval_accuracy":
0.5801526717557252, "test_eval_precision": 0.5666666666666667,
"test_eval_recall": 0.5396825396825397, "test_eval_f1": 0.5528455284552846,
"test_eval_runtime": 4.0852, "test_eval_samples_per_second": 224.47,
"test_eval_steps_per_second": 14.198, "test_epoch": 5.0}}
{"timestamp": "2025-04-09T19:58:58.496236-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 5,
"batch_size": 16, "weight_decay": 0.1, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.embeddings.", "bert.encoder.layer.0.", "bert.encoder.layer.8.",
"bert.encoder.layer.9.", "bert.encoder.layer.10.", "bert.encoder.layer.11.",
"bert.pooler.", "classifier."]}, "model_details": {"model_total_params":
108311810, "model_trainable_params": 58696706, "pooler_activation_function":
"Tanh", "config_attributes": {"return_dict": true, "output_hidden_states":
false, "output_attentions": false, "torchscript": false, "torch_dtype":
"float32", "use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":

```

```

1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.5292643904685974, "train_eval_accuracy": 0.7367528060558601,
"train_eval_precision": 0.7544311034877073, "train_eval_recall":
0.695022386094285, "train_eval_f1": 0.7235092529129541, "train_eval_runtime":
26.7206, "train_eval_samples_per_second": 286.745,
"train_eval_steps_per_second": 17.926, "train_epoch": 5.0, "val_eval_loss":
0.6574380993843079, "val_eval_accuracy": 0.6532066508313539,
"val_eval_precision": 0.6161616161616161, "val_eval_recall": 0.6354166666666666,
"val_eval_f1": 0.6256410256410256, "val_eval_runtime": 2.4237,
"val_eval_samples_per_second": 173.699, "val_eval_steps_per_second": 11.14,
"val_epoch": 5.0, "test_eval_loss": 0.7001335620880127, "test_eval_accuracy":
0.5910577971646674, "test_eval_precision": 0.5829145728643216,
"test_eval_recall": 0.5260770975056689, "test_eval_f1": 0.5530393325387366,
"test_eval_runtime": 4.2258, "test_eval_samples_per_second": 217.0,
"test_eval_steps_per_second": 13.725, "test_epoch": 5.0}}
{"timestamp": "2025-04-09T20:05:09.703123-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 5,
"batch_size": 16, "weight_decay": 0.1, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.embeddings.", "bert.encoder.layer.0.", "bert.encoder.layer.8.",
"bert.encoder.layer.9.", "bert.encoder.layer.10.", "bert.encoder.layer.11.",
"bert.pooler.", "classifier."]}, "model_details": {"model_total_params":
108311810, "model_trainable_params": 58696706, "pooler_activation_function":
"Tanh", "config_attributes": {"return_dict": true, "output_hidden_states":
false, "output_attentions": false, "torchscript": false, "torch_dtype":
"float32", "use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,

```

```

"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6857044696807861, "train_eval_accuracy": 0.5559906029757243,
"train_eval_precision": 0.5583801359739876, "train_eval_recall":
0.4974980247563866, "train_eval_f1": 0.526183844011142, "train_eval_runtime":
26.5746, "train_eval_samples_per_second": 288.321,
"train_eval_steps_per_second": 18.025, "train_epoch": 5.0, "val_eval_loss":
0.6887871623039246, "val_eval_accuracy": 0.5391923990498813,
"val_eval_precision": 0.4946808510638298, "val_eval_recall": 0.484375,
"val_eval_f1": 0.48947368421052634, "val_eval_runtime": 2.4568,
"val_eval_samples_per_second": 171.361, "val_eval_steps_per_second": 10.99,
"val_epoch": 5.0, "test_eval_loss": 0.6867599487304688, "test_eval_accuracy":
0.5321701199563795, "test_eval_precision": 0.5148514851485149,
"test_eval_recall": 0.47165532879818595, "test_eval_f1": 0.49230769230769234,
"test_eval_runtime": 4.1383, "test_eval_samples_per_second": 221.59,
"test_eval_steps_per_second": 14.015, "test_epoch": 5.0}}
{"timestamp": "2025-04-09T20:06:34.543851-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 5,
"batch_size": 16, "weight_decay": 0.1, "x_task": "multi", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.embeddings.", "bert.encoder.layer.0.", "bert.encoder.layer.8.",
"bert.encoder.layer.9.", "bert.encoder.layer.10.", "bert.encoder.layer.11.",
"bert.pooler.", "classifier."]}, "model_details": {"model_total_params":
108311810, "model_trainable_params": 58696706, "pooler_activation_function":
"Tanh", "config_attributes": {"return_dict": true, "output_hidden_states":

```

```

false, "output_attentions": false, "torchscript": false, "torch_dtype":
"float32", "use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.682285487651825, "train_eval_accuracy": 0.5655899802241265,
"train_eval_precision": 0.5964912280701754, "train_eval_recall":
0.40369393139841686, "train_eval_f1": 0.4815106215578285, "train_eval_runtime":
6.1019, "train_eval_samples_per_second": 248.611, "train_eval_steps_per_second":
15.569, "train_epoch": 5.0, "val_eval_loss": 0.7040739059448242,
"val_eval_accuracy": 0.48484848484848486, "val_eval_precision": 0.5,
"val_eval_recall": 0.3137254901960784, "val_eval_f1": 0.3855421686746988,
"val_eval_runtime": 1.7566, "val_eval_samples_per_second": 56.36,
"val_eval_steps_per_second": 3.985, "val_epoch": 5.0, "test_eval_loss":
0.69910728931427, "test_eval_accuracy": 0.4782608695652174,
"test_eval_precision": 0.5272727272727272, "test_eval_recall":
0.29292929292929293, "test_eval_f1": 0.37662337662337664, "test_eval_runtime":
1.9007, "test_eval_samples_per_second": 96.808, "test_eval_steps_per_second":
6.314, "test_epoch": 5.0}}
{"timestamp": "2025-04-09T20:12:45.781348-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 5,
"batch_size": 16, "weight_decay": 0.1, "x_task": "single", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.embeddings.", "bert.encoder.layer.0.", "bert.encoder.layer.8.",

```



```

"bert.encoder.layer.9.", "bert.encoder.layer.10.", "bert.encoder.layer.11.",
"bert.pooler.", "classifier."}], "model_details": {"model_total_params":
108311810, "model_trainable_params": 58696706, "pooler_activation_function":
"Tanh", "config_attributes": {"return_dict": true, "output_hidden_states":
false, "output_attentions": false, "torchscript": false, "torch_dtype":
"float32", "use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6809309720993042, "train_eval_accuracy": 0.5647350561211172,
"train_eval_precision": 0.5625, "train_eval_recall": 0.5475375296286542,
"train_eval_f1": 0.5549179233951689, "train_eval_runtime": 26.6489,
"train_eval_samples_per_second": 287.517, "train_eval_steps_per_second": 17.974,
"train_epoch": 5.0, "val_eval_loss": 0.6775543689727783, "val_eval_accuracy":
0.5866983372921615, "val_eval_precision": 0.5432692307692307, "val_eval_recall":
0.5885416666666666, "val_eval_f1": 0.565, "val_eval_runtime": 2.4459,
"val_eval_samples_per_second": 172.127, "val_eval_steps_per_second": 11.039,
"val_epoch": 5.0, "test_eval_loss": 0.6879650950431824, "test_eval_accuracy":
0.5463467829880043, "test_eval_precision": 0.5266524520255863,
"test_eval_recall": 0.5600907029478458, "test_eval_f1": 0.5428571428571428,
"test_eval_runtime": 4.077, "test_eval_samples_per_second": 224.923,
"test_eval_steps_per_second": 14.226, "test_epoch": 5.0}}
{"timestamp": "2025-04-09T20:14:12.327590-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 5,

```

```

"batch_size": 16, "weight_decay": 0.1, "x_task": "multi", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.embeddings.", "bert.encoder.layer.0.", "bert.encoder.layer.8.",
"bert.encoder.layer.9.", "bert.encoder.layer.10.", "bert.encoder.layer.11.",
"bert.pooler.", "classifier."], "model_details": {"model_total_params":
108311810, "model_trainable_params": 58696706, "pooler_activation_function":
"Tanh", "config_attributes": {"return_dict": true, "output_hidden_states":
false, "output_attentions": false, "torchscript": false, "torch_dtype":
"float32", "use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6613212823867798, "train_eval_accuracy": 0.6104152933421226,
"train_eval_precision": 0.6100131752305665, "train_eval_recall":
0.6108179419525066, "train_eval_f1": 0.6104152933421226, "train_eval_runtime":
6.2418, "train_eval_samples_per_second": 243.038, "train_eval_steps_per_second":
15.22, "train_epoch": 5.0, "val_eval_loss": 0.6623957753181458,
"val_eval_accuracy": 0.6161616161616161, "val_eval_precision":
0.6326530612244898, "val_eval_recall": 0.6078431372549019, "val_eval_f1": 0.62,
"val_eval_runtime": 1.7928, "val_eval_samples_per_second": 55.222,
"val_eval_steps_per_second": 3.905, "val_epoch": 5.0, "test_eval_loss":
0.7039431929588318, "test_eval_accuracy": 0.5217391304347826,
"test_eval_precision": 0.5679012345679012, "test_eval_recall":
0.46464646464646464, "test_eval_f1": 0.5111111111111111, "test_eval_runtime":

```

```

2.0408, "test_eval_samples_per_second": 90.159, "test_eval_steps_per_second":
5.88, "test_epoch": 5.0}}
{"timestamp": "2025-04-09T20:20:23.335352-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 5,
"batch_size": 16, "weight_decay": 0.1, "x_task": "single", "x_col": "sentence",
"y_col": "binary_complexity", "layers_to_unfreeze": ["bert.embeddings.",
"bert.encoder.layer.0.", "bert.encoder.layer.8.", "bert.encoder.layer.9.",
"bert.encoder.layer.10.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."], "model_details": {"model_total_params": 108311810,
"model_trainable_params": 58696706, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.5323772430419922, "train_eval_accuracy": 0.7358392064735056,
"train_eval_precision": 0.7576285963382737, "train_eval_recall":
0.6865946800105346, "train_eval_f1": 0.7203647416413373, "train_eval_runtime":
26.7046, "train_eval_samples_per_second": 286.917,
"train_eval_steps_per_second": 17.937, "train_epoch": 5.0, "val_eval_loss":
0.6548150181770325, "val_eval_accuracy": 0.6579572446555819,
"val_eval_precision": 0.6237113402061856, "val_eval_recall": 0.6302083333333334,
"val_eval_f1": 0.6269430051813472, "val_eval_runtime": 2.4259,

```

```

"val_eval_samples_per_second": 173.541, "val_eval_steps_per_second": 11.13,
"val_epoch": 5.0, "test_eval_loss": 0.6973928809165955, "test_eval_accuracy":
0.5921483097055616, "test_eval_precision": 0.5856777493606138,
"test_eval_recall": 0.5192743764172335, "test_eval_f1": 0.5504807692307693,
"test_eval_runtime": 4.3909, "test_eval_samples_per_second": 208.841,
"test_eval_steps_per_second": 13.209, "test_epoch": 5.0}}
{"timestamp": "2025-04-09T20:21:50.237171-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 5,
"batch_size": 16, "weight_decay": 0.1, "x_task": "multi", "x_col": "sentence",
"y_col": "binary_complexity", "layers_to_unfreeze": ["bert.embeddings.",
"bert.encoder.layer.0.", "bert.encoder.layer.8.", "bert.encoder.layer.9.",
"bert.encoder.layer.10.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 58696706, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.5702769756317139, "train_eval_accuracy": 0.7066578773895847,
"train_eval_precision": 0.7809694793536804, "train_eval_recall":
0.5738786279683378, "train_eval_f1": 0.6615969581749049, "train_eval_runtime":
6.2164, "train_eval_samples_per_second": 244.031, "train_eval_steps_per_second":

```

```

15.282, "train_epoch": 5.0, "val_eval_loss": 0.7027019262313843,
"val_eval_accuracy": 0.5353535353535354, "val_eval_precision":
0.5675675675675675, "val_eval_recall": 0.4117647058823529, "val_eval_f1":
0.4772727272727273, "val_eval_runtime": 2.0135, "val_eval_samples_per_second":
49.168, "val_eval_steps_per_second": 3.477, "val_epoch": 5.0, "test_eval_loss":
0.6995925307273865, "test_eval_accuracy": 0.5489130434782609,
"test_eval_precision": 0.6428571428571429, "test_eval_recall":
0.36363636363636365, "test_eval_f1": 0.4645161290322581, "test_eval_runtime":
2.5655, "test_eval_samples_per_second": 71.722, "test_eval_steps_per_second":
4.678, "test_epoch": 5.0}}
{"timestamp": "2025-04-09T20:44:10.201866-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 5,
"batch_size": 8, "weight_decay": 0, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.embeddings.", "bert.encoder.layer.0.", "bert.encoder.layer.1.",
"bert.encoder.layer.8.", "bert.encoder.layer.9.", "bert.encoder.layer.10.",
"bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 65784578,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":

```

```

0.4137938320636749, "train_eval_accuracy": 0.8181936831114591,
"train_eval_precision": 0.8605278944211158, "train_eval_recall":
0.7555965235712404, "train_eval_f1": 0.8046557285093254, "train_eval_runtime":
21.3952, "train_eval_samples_per_second": 358.118,
"train_eval_steps_per_second": 44.776, "train_epoch": 5.0, "val_eval_loss":
0.7340455651283264, "val_eval_accuracy": 0.6342042755344418,
"val_eval_precision": 0.6091954022988506, "val_eval_recall": 0.5520833333333334,
"val_eval_f1": 0.5792349726775956, "val_eval_runtime": 6.0524,
"val_eval_samples_per_second": 69.559, "val_eval_steps_per_second": 8.757,
"val_epoch": 5.0, "test_eval_loss": 0.811907947063446, "test_eval_accuracy":
0.5997818974918212, "test_eval_precision": 0.6010928961748634,
"test_eval_recall": 0.4988662131519274, "test_eval_f1": 0.5452292441140025,
"test_eval_runtime": 6.815, "test_eval_samples_per_second": 134.556,
"test_eval_steps_per_second": 16.875, "test_epoch": 5.0}}
{"timestamp": "2025-04-09T20:48:58.823946-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 5,
"batch_size": 8, "weight_decay": 0, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.embeddings.", "bert.encoder.layer.0.", "bert.encoder.layer.1.",
"bert.encoder.layer.8.", "bert.encoder.layer.9.", "bert.encoder.layer.10.",
"bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 65784578,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,

```

```

"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.4213423728942871, "train_eval_accuracy": 0.8101018010963195,
"train_eval_precision": 0.8272778088317496, "train_eval_recall":
0.7795628127469054, "train_eval_f1": 0.8027118644067797, "train_eval_runtime":
21.1208, "train_eval_samples_per_second": 362.77, "train_eval_steps_per_second":
45.358, "train_epoch": 5.0, "val_eval_loss": 0.7140721678733826,
"val_eval_accuracy": 0.6270783847980997, "val_eval_precision":
0.5888324873096447, "val_eval_recall": 0.6041666666666666, "val_eval_f1":
0.596401028277635, "val_eval_runtime": 5.6721, "val_eval_samples_per_second":
74.223, "val_eval_steps_per_second": 9.344, "val_epoch": 5.0, "test_eval_loss":
0.7822856903076172, "test_eval_accuracy": 0.6074154852780806,
"test_eval_precision": 0.6015037593984962, "test_eval_recall": 0.54421768707483,
"test_eval_f1": 0.5714285714285714, "test_eval_runtime": 7.0926,
"test_eval_samples_per_second": 129.289, "test_eval_steps_per_second": 16.214,
"test_epoch": 5.0}}
{"timestamp": "2025-04-09T20:53:49.509607-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 5,
"batch_size": 8, "weight_decay": 0, "x_task": "single", "x_col": "pos_sequence",
"y_col": "binary_complexity", "layers_to_unfreeze": ["bert.embeddings.",
"bert.encoder.layer.0.", "bert.encoder.layer.1.", "bert.encoder.layer.8.",
"bert.encoder.layer.9.", "bert.encoder.layer.10.", "bert.encoder.layer.11.",
"bert.pooler.", "classifier."]}, "model_details": {"model_total_params":
108311810, "model_trainable_params": 65784578, "pooler_activation_function":
"Tanh", "config_attributes": {"return_dict": true, "output_hidden_states":
false, "output_attentions": false, "torchscript": false, "torch_dtype":
"float32", "use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":

```

```

"sdpa", "attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6846482157707214, "train_eval_accuracy": 0.5565126598799269,
"train_eval_precision": 0.5684391080617496, "train_eval_recall":
0.43639715564919673, "train_eval_f1": 0.4937425506555423, "train_eval_runtime":
21.3677, "train_eval_samples_per_second": 358.578,
"train_eval_steps_per_second": 44.834, "train_epoch": 5.0, "val_eval_loss":
0.6823729276657104, "val_eval_accuracy": 0.5605700712589073,
"val_eval_precision": 0.5202312138728323, "val_eval_recall": 0.46875,
"val_eval_f1": 0.4931506849315068, "val_eval_runtime": 5.7756,
"val_eval_samples_per_second": 72.893, "val_eval_steps_per_second": 9.177,
"val_epoch": 5.0, "test_eval_loss": 0.6859716773033142, "test_eval_accuracy":
0.5528898582333697, "test_eval_precision": 0.5439093484419264,
"test_eval_recall": 0.43537414965986393, "test_eval_f1": 0.4836272040302267,
"test_eval_runtime": 6.7389, "test_eval_samples_per_second": 136.076,
"test_eval_steps_per_second": 17.065, "test_epoch": 5.0}}
{"timestamp": "2025-04-09T20:55:30.176738-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 5,
"batch_size": 8, "weight_decay": 0, "x_task": "multi", "x_col": "pos_sequence",
"y_col": "binary_complexity", "layers_to_unfreeze": ["bert.embeddings.",
"bert.encoder.layer.0.", "bert.encoder.layer.1.", "bert.encoder.layer.8.",
"bert.encoder.layer.9.", "bert.encoder.layer.10.", "bert.encoder.layer.11.",
"bert.pooler.", "classifier."], "model_details": {"model_total_params":
108311810, "model_trainable_params": 65784578, "pooler_activation_function":
"Tanh", "config_attributes": {"return_dict": true, "output_hidden_states":
false, "output_attentions": false, "torchscript": false, "torch_dtype":
"float32", "use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":

```



```

0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdp", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6737680435180664, "train_eval_accuracy": 0.5774555042847725,
"train_eval_precision": 0.5677867902665121, "train_eval_recall":
0.6464379947229552, "train_eval_f1": 0.6045650832819247, "train_eval_runtime":
8.5349, "train_eval_samples_per_second": 177.741, "train_eval_steps_per_second":
22.262, "train_epoch": 5.0, "val_eval_loss": 0.7262722253799438,
"val_eval_accuracy": 0.48484848484848486, "val_eval_precision": 0.5,
"val_eval_recall": 0.5490196078431373, "val_eval_f1": 0.5233644859813084,
"val_eval_runtime": 5.1128, "val_eval_samples_per_second": 19.363,
"val_eval_steps_per_second": 2.543, "val_epoch": 5.0, "test_eval_loss":
0.7065175771713257, "test_eval_accuracy": 0.5108695652173914,
"test_eval_precision": 0.5517241379310345, "test_eval_recall":
0.48484848484848486, "test_eval_f1": 0.5161290322580645, "test_eval_runtime":
5.3479, "test_eval_samples_per_second": 34.406, "test_eval_steps_per_second":
4.301, "test_epoch": 5.0}}
{"timestamp": "2025-04-09T21:00:15.721885-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 5,
"batch_size": 8, "weight_decay": 0, "x_task": "single", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.embeddings.", "bert.encoder.layer.0.", "bert.encoder.layer.8.",
"bert.encoder.layer.9.", "bert.encoder.layer.10.", "bert.encoder.layer.11.",
"bert.pooler.", "classifier."]}, "model_details": {"model_total_params":
108311810, "model_trainable_params": 58696706, "pooler_activation_function":
"Tanh", "config_attributes": {"return_dict": true, "output_hidden_states":
false, "output_attentions": false, "torchscript": false, "torch_dtype":
"float32", "use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,

```

```

"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6860324144363403, "train_eval_accuracy": 0.5395458104933438,
"train_eval_precision": 0.5453014482990906, "train_eval_recall":
0.4263892546747432, "train_eval_f1": 0.47856931717410583, "train_eval_runtime":
21.7179, "train_eval_samples_per_second": 352.796,
"train_eval_steps_per_second": 44.111, "train_epoch": 5.0, "val_eval_loss":
0.677761435508728, "val_eval_accuracy": 0.5344418052256532,
"val_eval_precision": 0.48823529411764705, "val_eval_recall":
0.4322916666666667, "val_eval_f1": 0.4585635359116022, "val_eval_runtime":
5.6131, "val_eval_samples_per_second": 75.004, "val_eval_steps_per_second":
9.442, "val_epoch": 5.0, "test_eval_loss": 0.6860611438751221,
"test_eval_accuracy": 0.5583424209378408, "test_eval_precision":
0.5494505494505495, "test_eval_recall": 0.45351473922902497, "test_eval_f1":
0.4968944099378882, "test_eval_runtime": 7.2528, "test_eval_samples_per_second":
126.434, "test_eval_steps_per_second": 15.856, "test_epoch": 5.0}}
{"timestamp": "2025-04-09T21:02:00.254676-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 1e-05, "epochs": 5,
"batch_size": 8, "weight_decay": 0, "x_task": "multi", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.embeddings.", "bert.encoder.layer.0.", "bert.encoder.layer.1.",
"bert.encoder.layer.8.", "bert.encoder.layer.9.", "bert.encoder.layer.10.",
"bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 65784578,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,

```

```

"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6459857225418091, "train_eval_accuracy": 0.6407382992748847,
"train_eval_precision": 0.6287787182587666, "train_eval_recall":
0.6860158311345647, "train_eval_f1": 0.6561514195583596, "train_eval_runtime":
8.1972, "train_eval_samples_per_second": 185.064, "train_eval_steps_per_second":
23.179, "train_epoch": 5.0, "val_eval_loss": 0.7082846760749817,
"val_eval_accuracy": 0.5050505050505051, "val_eval_precision":
0.5178571428571429, "val_eval_recall": 0.5686274509803921, "val_eval_f1":
0.5420560747663551, "val_eval_runtime": 5.5427, "val_eval_samples_per_second":
17.861, "val_eval_steps_per_second": 2.345, "val_epoch": 5.0, "test_eval_loss":
0.7356353402137756, "test_eval_accuracy": 0.4945652173913043,
"test_eval_precision": 0.5340909090909091, "test_eval_recall":
0.47474747474747475, "test_eval_f1": 0.5026737967914439, "test_eval_runtime":
5.1728, "test_eval_samples_per_second": 35.571, "test_eval_steps_per_second":
4.446, "test_epoch": 5.0}}
{"timestamp": "2025-04-10T10:38:14.118560-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 2e-05, "epochs": 2,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.8.", "bert.encoder.layer.9.", "bert.encoder.layer.10.",
"bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 28943618,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":

```

```

1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"eager", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6934679746627808, "train_eval_accuracy": 0.5069172539806839,
"train_eval_precision": 0.5049146404552509, "train_eval_recall":
0.257045035554385, "train_eval_f1": 0.3406631762652705, "train_eval_runtime":
13.0438, "train_eval_samples_per_second": 588.787,
"train_eval_steps_per_second": 4.6, "train_epoch": 2.0, "val_eval_loss":
0.6967368125915527, "val_eval_accuracy": 0.5415676959619953,
"val_eval_precision": 0.49473684210526314, "val_eval_recall":
0.24479166666666666, "val_eval_f1": 0.32752613240418116, "val_eval_runtime":
1.2098, "val_eval_samples_per_second": 423.219, "val_eval_steps_per_second":
3.306, "val_epoch": 2.0, "test_eval_loss": 0.6922365427017212,
"test_eval_accuracy": 0.5398037077426391, "test_eval_precision":
0.5381526104417671, "test_eval_recall": 0.30385487528344673, "test_eval_f1":
0.3884057971014493, "test_eval_runtime": 1.3965, "test_eval_samples_per_second":
733.25, "test_eval_steps_per_second": 5.729, "test_epoch": 2.0}}
{"timestamp": "2025-04-10T10:39:22.769353-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 2,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.8.", "bert.encoder.layer.9.", "bert.encoder.layer.10.",
"bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 28943618,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,

```

```

"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"eager", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6931834816932678, "train_eval_accuracy": 0.5156617071260767,
"train_eval_precision": 0.5196167883211679, "train_eval_recall":
0.2999736634184883, "train_eval_f1": 0.3803640006678911, "train_eval_runtime":
4.7831, "train_eval_samples_per_second": 1605.664,
"train_eval_steps_per_second": 12.544, "train_epoch": 2.0, "val_eval_loss":
0.6993261575698853, "val_eval_accuracy": 0.5296912114014252,
"val_eval_precision": 0.47619047619047616, "val_eval_recall": 0.3125,
"val_eval_f1": 0.37735849056603776, "val_eval_runtime": 1.942,
"val_eval_samples_per_second": 263.64, "val_eval_steps_per_second": 2.06,
"val_epoch": 2.0, "test_eval_loss": 0.6964348554611206, "test_eval_accuracy":
0.495092693565976, "test_eval_precision": 0.4566929133858268,
"test_eval_recall": 0.26303854875283444, "test_eval_f1": 0.3338129496402878,
"test_eval_runtime": 1.4322, "test_eval_samples_per_second": 714.989,
"test_eval_steps_per_second": 5.586, "test_epoch": 2.0}}
{"timestamp": "2025-04-10T10:40:35.651600-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 5,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.8.", "bert.encoder.layer.9.", "bert.encoder.layer.10.",
"bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 28943618,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,

```

```

"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"eager", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6920309662818909, "train_eval_accuracy": 0.5200991908117985,
"train_eval_precision": 0.5202020202020202, "train_eval_recall":
0.40690018435607056, "train_eval_f1": 0.45662775232747155, "train_eval_runtime":
4.2772, "train_eval_samples_per_second": 1795.573,
"train_eval_steps_per_second": 14.028, "train_epoch": 5.0, "val_eval_loss":
0.696632981300354, "val_eval_accuracy": 0.5083135391923991,
"val_eval_precision": 0.4556213017751479, "val_eval_recall": 0.4010416666666667,
"val_eval_f1": 0.4265927977839335, "val_eval_runtime": 1.319,
"val_eval_samples_per_second": 388.182, "val_eval_steps_per_second": 3.033,
"val_epoch": 5.0, "test_eval_loss": 0.6927960515022278, "test_eval_accuracy":
0.520174482006543, "test_eval_precision": 0.5014577259475219,
"test_eval_recall": 0.3900226757369615, "test_eval_f1": 0.4387755102040816,
"test_eval_runtime": 1.9242, "test_eval_samples_per_second": 532.162,
"test_eval_steps_per_second": 4.158, "test_epoch": 5.0}}
{"timestamp": "2025-04-10T10:44:05.039614-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 3,
"batch_size": 8, "weight_decay": 0.01, "x_task": "multi", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.8.", "bert.encoder.layer.9.", "bert.encoder.layer.10.",

```

```

"bert.encoder.layer.11.", "bert.pooler.", "classifier."]], "model_details":
{"model_total_params": 108311810, "model_trainable_params": 28943618,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"eager", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.691429078578949, "train_eval_accuracy": 0.5102175346077785,
"train_eval_precision": 0.5051405071967101, "train_eval_recall":
0.9722955145118733, "train_eval_f1": 0.6648624267027514, "train_eval_runtime":
20.6976, "train_eval_samples_per_second": 73.439, "train_eval_steps_per_second":
9.18, "train_epoch": 3.0, "val_eval_loss": 0.6856585144996643,
"val_eval_accuracy": 0.5555555555555556, "val_eval_precision":
0.5368421052631579, "val_eval_recall": 1.0, "val_eval_f1": 0.6986301369863014,
"val_eval_runtime": 1.3199, "val_eval_samples_per_second": 78.793,
"val_eval_steps_per_second": 9.849, "val_epoch": 3.0, "test_eval_loss":
0.6929100751876831, "test_eval_accuracy": 0.5163043478260869,
"test_eval_precision": 0.5277777777777778, "test_eval_recall":
0.9595959595959596, "test_eval_f1": 0.6810035842293907, "test_eval_runtime":
1.5264, "test_eval_samples_per_second": 120.546, "test_eval_steps_per_second":
15.068, "test_epoch": 3.0}}
{"timestamp": "2025-04-10T14:02:50.730999-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,

```

```

"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.10.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]], "model_details": {"model_total_params": 108311810,
"model_trainable_params": 14767874, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6900879740715027, "train_eval_accuracy": 0.5211433046202036,
"train_eval_precision": 0.509794918885828, "train_eval_recall":
0.8772715301553858, "train_eval_f1": 0.644855289904172, "train_eval_runtime":
18.8806, "train_eval_samples_per_second": 405.813,
"train_eval_steps_per_second": 3.178, "train_epoch": 1.0, "val_eval_loss":
0.6939961314201355, "val_eval_accuracy": 0.505938242280285,
"val_eval_precision": 0.4772727272727273, "val_eval_recall": 0.875,
"val_eval_f1": 0.6176470588235294, "val_eval_runtime": 5.3727,
"val_eval_samples_per_second": 78.359, "val_eval_steps_per_second": 0.745,
"val_epoch": 1.0, "test_eval_loss": 0.6909573674201965, "test_eval_accuracy":
0.4994547437295529, "test_eval_precision": 0.4886934673366834,
"test_eval_recall": 0.8820861678004536, "test_eval_f1": 0.6289409862570736,
"test_eval_runtime": 6.0905, "test_eval_samples_per_second": 150.562,

```



```

"test_eval_steps_per_second": 1.314, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T14:23:32.888277-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.10.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 14767874, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.689695417881012, "train_eval_accuracy": 0.538371182458888,
"train_eval_precision": 0.5458069062720226, "train_eval_recall":
0.40795364761653935, "train_eval_f1": 0.4669178598342125, "train_eval_runtime":
16.1142, "train_eval_samples_per_second": 475.481,
"train_eval_steps_per_second": 3.723, "train_epoch": 1.0, "val_eval_loss":
0.6870003938674927, "val_eval_accuracy": 0.5558194774346793,
"val_eval_precision": 0.515527950310559, "val_eval_recall": 0.4322916666666667,
"val_eval_f1": 0.4702549575070821, "val_eval_runtime": 5.3788,
"val_eval_samples_per_second": 78.271, "val_eval_steps_per_second": 0.744,
"val_epoch": 1.0, "test_eval_loss": 0.6854057908058167, "test_eval_accuracy":

```

```

0.5561613958560524, "test_eval_precision": 0.5578231292517006,
"test_eval_recall": 0.37188208616780044, "test_eval_f1": 0.44625850340136053,
"test_eval_runtime": 6.1126, "test_eval_samples_per_second": 150.019,
"test_eval_steps_per_second": 1.309, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T14:25:44.268310-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7430146336555481, "train_eval_accuracy": 0.4958235447663795,
"train_eval_precision": 0.49569190600522195, "train_eval_recall": 1.0,
"train_eval_f1": 0.6628262197783015, "train_eval_runtime": 16.7201,
"train_eval_samples_per_second": 458.25, "train_eval_steps_per_second": 3.588,
"train_epoch": 1.0, "val_eval_loss": 0.7697564959526062, "val_eval_accuracy":
0.45605700712589076, "val_eval_precision": 0.45605700712589076,
"val_eval_recall": 1.0, "val_eval_f1": 0.6264274061990212, "val_eval_runtime":
5.6903, "val_eval_samples_per_second": 73.985, "val_eval_steps_per_second":

```

```

0.703, "val_epoch": 1.0, "test_eval_loss": 0.7486463189125061,
"test_eval_accuracy": 0.48091603053435117, "test_eval_precision":
0.48091603053435117, "test_eval_recall": 1.0, "test_eval_f1":
0.6494845360824743, "test_eval_runtime": 6.1595, "test_eval_samples_per_second":
148.875, "test_eval_steps_per_second": 1.299, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T14:32:42.886741-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.10.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 14767874, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.708980143070221, "train_eval_accuracy": 0.49569303054032887,
"train_eval_precision": 0.49562720271505023, "train_eval_recall": 1.0,
"train_eval_f1": 0.662768371443533, "train_eval_runtime": 16.5925,
"train_eval_samples_per_second": 461.775, "train_eval_steps_per_second": 3.616,
"train_epoch": 1.0, "val_eval_loss": 0.7232958674430847, "val_eval_accuracy":
0.45605700712589076, "val_eval_precision": 0.45605700712589076,

```

```

"val_eval_recall": 1.0, "val_eval_f1": 0.6264274061990212, "val_eval_runtime":
6.1138, "val_eval_samples_per_second": 68.861, "val_eval_steps_per_second":
0.654, "val_epoch": 1.0, "test_eval_loss": 0.7155287861824036,
"test_eval_accuracy": 0.48091603053435117, "test_eval_precision":
0.48091603053435117, "test_eval_recall": 1.0, "test_eval_f1":
0.6494845360824743, "test_eval_runtime": 6.077, "test_eval_samples_per_second":
150.898, "test_eval_steps_per_second": 1.316, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T14:38:01.905420-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 3,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.10.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 14767874, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null, "gradient_checkpointing": false}}, "run_metrics": {"train_eval_loss":
0.6901301741600037, "train_eval_accuracy": 0.5371965544244323,
"train_eval_precision": 0.5365144020948501, "train_eval_recall":
0.4856465630761127, "train_eval_f1": 0.5098147636162565, "train_eval_runtime":
16.1129, "train_eval_samples_per_second": 475.521,

```

```

"train_eval_steps_per_second": 3.724, "train_epoch": 3.0, "val_eval_loss":
0.6896899938583374, "val_eval_accuracy": 0.5201900237529691,
"val_eval_precision": 0.474747474747475, "val_eval_recall":
0.4895833333333333, "val_eval_f1": 0.48205128205128206, "val_eval_runtime":
5.7523, "val_eval_samples_per_second": 73.188, "val_eval_steps_per_second":
0.695, "val_epoch": 3.0, "test_eval_loss": 0.6892867088317871,
"test_eval_accuracy": 0.5387131952017448, "test_eval_precision":
0.5210280373831776, "test_eval_recall": 0.5056689342403629, "test_eval_f1":
0.5132336018411968, "test_eval_runtime": 6.5784, "test_eval_samples_per_second":
139.395, "test_eval_steps_per_second": 1.216, "test_epoch": 3.0}}
{"timestamp": "2025-04-10T14:46:15.932379-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.10.", "bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 14767874, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6888505220413208, "train_eval_accuracy": 0.5390237535891412,

```

```

"train_eval_precision": 0.5320280396422529, "train_eval_recall":
0.5796681590729523, "train_eval_f1": 0.5548273254348374, "train_eval_runtime":
16.2993, "train_eval_samples_per_second": 470.082,
"train_eval_steps_per_second": 3.681, "train_epoch": 1.0, "val_eval_loss":
0.686075747013092, "val_eval_accuracy": 0.5700712589073634,
"val_eval_precision": 0.5244444444444445, "val_eval_recall": 0.6145833333333334,
"val_eval_f1": 0.565947242206235, "val_eval_runtime": 6.2542,
"val_eval_samples_per_second": 67.315, "val_eval_steps_per_second": 0.64,
"val_epoch": 1.0, "test_eval_loss": 0.6869158744812012, "test_eval_accuracy":
0.5561613958560524, "test_eval_precision": 0.5358649789029536,
"test_eval_recall": 0.5759637188208617, "test_eval_f1": 0.5551912568306011,
"test_eval_runtime": 6.0502, "test_eval_samples_per_second": 151.565,
"test_eval_steps_per_second": 1.322, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T14:47:28.456414-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":

```

```

true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7430146336555481, "train_eval_accuracy": 0.4958235447663795,
"train_eval_precision": 0.49569190600522195, "train_eval_recall": 1.0,
"train_eval_f1": 0.6628262197783015, "train_eval_runtime": 16.954,
"train_eval_samples_per_second": 451.928, "train_eval_steps_per_second": 3.539,
"train_epoch": 1.0, "val_eval_loss": 0.7697564959526062, "val_eval_accuracy":
0.45605700712589076, "val_eval_precision": 0.45605700712589076,
"val_eval_recall": 1.0, "val_eval_f1": 0.6264274061990212, "val_eval_runtime":
5.3997, "val_eval_samples_per_second": 77.968, "val_eval_steps_per_second":
0.741, "val_epoch": 1.0, "test_eval_loss": 0.7486463189125061,
"test_eval_accuracy": 0.48091603053435117, "test_eval_precision":
0.48091603053435117, "test_eval_recall": 1.0, "test_eval_f1":
0.6494845360824743, "test_eval_runtime": 6.0762, "test_eval_samples_per_second":
150.918, "test_eval_steps_per_second": 1.317, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T14:48:34.667878-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 124647170, "model_trainable_params": 592130,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffb7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,

```

```

"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6939288377761841,
"train_eval_accuracy": 0.4955625163142783, "train_eval_precision":
0.4955625163142783, "train_eval_recall": 1.0, "train_eval_f1":
0.6627105332053408, "train_eval_runtime": 16.2247,
"train_eval_samples_per_second": 472.244, "train_eval_steps_per_second": 3.698,
"train_epoch": 1.0, "val_eval_loss": 0.6955541968345642, "val_eval_accuracy":
0.45605700712589076, "val_eval_precision": 0.45605700712589076,
"val_eval_recall": 1.0, "val_eval_f1": 0.6264274061990212, "val_eval_runtime":
5.3345, "val_eval_samples_per_second": 78.921, "val_eval_steps_per_second":
0.75, "val_epoch": 1.0, "test_eval_loss": 0.694735586643219,
"test_eval_accuracy": 0.48091603053435117, "test_eval_precision":
0.48091603053435117, "test_eval_recall": 1.0, "test_eval_f1":
0.6494845360824743, "test_eval_runtime": 6.1612, "test_eval_samples_per_second":
148.835, "test_eval_steps_per_second": 1.298, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T14:58:24.702471-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaeff65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,

```



```

"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.7858944535255432,
"train_eval_accuracy": 0.49895588619159487, "train_eval_precision":
0.496943231441048, "train_eval_recall": 0.8991308928101133, "train_eval_f1":
0.6401049967188526, "train_eval_runtime": 39.7282,
"train_eval_samples_per_second": 192.86, "train_eval_steps_per_second": 1.51,
"train_epoch": 1.0, "val_eval_loss": 0.8127373456954956, "val_eval_accuracy":
0.4679334916864608, "val_eval_precision": 0.45675675675675675,
"val_eval_recall": 0.8802083333333334, "val_eval_f1": 0.6014234875444839,
"val_eval_runtime": 6.6549, "val_eval_samples_per_second": 63.261,
"val_eval_steps_per_second": 0.601, "val_epoch": 1.0, "test_eval_loss":
0.7970331311225891, "test_eval_accuracy": 0.48745910577971646,
"test_eval_precision": 0.4825090470446321, "test_eval_recall":
0.9070294784580499, "test_eval_f1": 0.6299212598425197, "test_eval_runtime":
8.9081, "test_eval_samples_per_second": 102.94, "test_eval_steps_per_second":
0.898, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T15:11:57.683084-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,

```

```

"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.7042792439460754,
"train_eval_accuracy": 0.5074393108848865, "train_eval_precision":
0.5022642252411892, "train_eval_recall": 0.6718461943639715, "train_eval_f1":
0.574808472284813, "train_eval_runtime": 39.6698,
"train_eval_samples_per_second": 193.144, "train_eval_steps_per_second": 1.512,
"train_epoch": 1.0, "val_eval_loss": 0.7087881565093994, "val_eval_accuracy":
0.4916864608076009, "val_eval_precision": 0.460431654676259, "val_eval_recall":
0.6666666666666666, "val_eval_f1": 0.5446808510638298, "val_eval_runtime":
6.6967, "val_eval_samples_per_second": 62.866, "val_eval_steps_per_second":
0.597, "val_epoch": 1.0, "test_eval_loss": 0.7028823494911194,
"test_eval_accuracy": 0.5005452562704471, "test_eval_precision":
0.48500881834215165, "test_eval_recall": 0.6235827664399093, "test_eval_f1":
0.5456349206349206, "test_eval_runtime": 9.502, "test_eval_samples_per_second":
96.506, "test_eval_steps_per_second": 0.842, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T15:20:13.862382-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",

```

```

"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6927051544189453,
"train_eval_accuracy": 0.5135734795092665, "train_eval_precision":
0.5050754060324826, "train_eval_recall": 0.9173031340531999, "train_eval_f1":
0.6514542223884785, "train_eval_runtime": 16.0671,
"train_eval_samples_per_second": 476.875, "train_eval_steps_per_second": 3.734,
"train_epoch": 1.0, "val_eval_loss": 0.6937984228134155, "val_eval_accuracy":

```

```

0.4750593824228028, "val_eval_precision": 0.4609164420485175, "val_eval_recall":
0.890625, "val_eval_f1": 0.6074600355239786, "val_eval_runtime": 5.9955,
"val_eval_samples_per_second": 70.219, "val_eval_steps_per_second": 0.667,
"val_epoch": 1.0, "test_eval_loss": 0.6928855776786804, "test_eval_accuracy":
0.48854961832061067, "test_eval_precision": 0.4828009828009828,
"test_eval_recall": 0.891156462585034, "test_eval_f1": 0.6262948207171315,
"test_eval_runtime": 6.141, "test_eval_samples_per_second": 149.324,
"test_eval_steps_per_second": 1.303, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T15:25:30.507330-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 2,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaeff65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6932896375656128,
"train_eval_accuracy": 0.5097885669537979, "train_eval_precision":

```

```

0.5039129604886429, "train_eval_recall": 0.6952857519094021, "train_eval_f1":
0.5843293492695883, "train_eval_runtime": 39.9183,
"train_eval_samples_per_second": 191.942, "train_eval_steps_per_second": 1.503,
"train_epoch": 2.0, "val_eval_loss": 0.6949265599250793, "val_eval_accuracy":
0.4845605700712589, "val_eval_precision": 0.4551971326164875, "val_eval_recall":
0.6614583333333334, "val_eval_f1": 0.5392781316348195, "val_eval_runtime":
6.7622, "val_eval_samples_per_second": 62.258, "val_eval_steps_per_second":
0.592, "val_epoch": 2.0, "test_eval_loss": 0.6902783513069153,
"test_eval_accuracy": 0.5223555070883316, "test_eval_precision":
0.5022900763358779, "test_eval_recall": 0.746031746031746, "test_eval_f1":
0.6003649635036497, "test_eval_runtime": 8.9678, "test_eval_samples_per_second":
102.255, "test_eval_steps_per_second": 0.892, "test_epoch": 2.0}}
{"timestamp": "2025-04-10T15:27:54.747459-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,

```

```

"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6941861510276794,
"train_eval_accuracy": 0.5044374836857217, "train_eval_precision": 0.0,
"train_eval_recall": 0.0, "train_eval_f1": 0.0, "train_eval_runtime": 40.2534,
"train_eval_samples_per_second": 190.344, "train_eval_steps_per_second": 1.491,
"train_epoch": 1.0, "val_eval_loss": 0.6885123252868652, "val_eval_accuracy":
0.5439429928741093, "val_eval_precision": 0.0, "val_eval_recall": 0.0,
"val_eval_f1": 0.0, "val_eval_runtime": 6.6585, "val_eval_samples_per_second":
63.227, "val_eval_steps_per_second": 0.601, "val_epoch": 1.0, "test_eval_loss":
0.6924662590026855, "test_eval_accuracy": 0.5190839694656488,
"test_eval_precision": 0.0, "test_eval_recall": 0.0, "test_eval_f1": 0.0,
"test_eval_runtime": 8.9276, "test_eval_samples_per_second": 102.715,
"test_eval_steps_per_second": 0.896, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T15:31:32.436306-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 3,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",

```

```

"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6893024444580078,
"train_eval_accuracy": 0.5486818063168886, "train_eval_precision":
0.5789473684210527, "train_eval_recall": 0.32736370819067684, "train_eval_f1":
0.41823687752355315, "train_eval_runtime": 39.6203,
"train_eval_samples_per_second": 193.386, "train_eval_steps_per_second": 1.514,
"train_epoch": 3.0, "val_eval_loss": 0.6867628693580627, "val_eval_accuracy":
0.5843230403800475, "val_eval_precision": 0.5739130434782609, "val_eval_recall":
0.34375, "val_eval_f1": 0.42996742671009774, "val_eval_runtime": 6.8568,
"val_eval_samples_per_second": 61.399, "val_eval_steps_per_second": 0.583,
"val_epoch": 3.0, "test_eval_loss": 0.6874121427536011, "test_eval_accuracy":
0.5507088331515813, "test_eval_precision": 0.5627705627705628,
"test_eval_recall": 0.2947845804988662, "test_eval_f1": 0.3869047619047619,
"test_eval_runtime": 8.9906, "test_eval_samples_per_second": 101.996,
"test_eval_steps_per_second": 0.89, "test_epoch": 3.0}}
{"timestamp": "2025-04-10T15:34:19.156321-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 2,

```

```

"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,

```



```

"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6899241805076599,
"train_eval_accuracy": 0.5409814669799008, "train_eval_precision":
0.5381887615930169, "train_eval_recall": 0.5196207532262312, "train_eval_f1":
0.5287417928447006, "train_eval_runtime": 39.3313,
"train_eval_samples_per_second": 194.807, "train_eval_steps_per_second": 1.526,
"train_epoch": 2.0, "val_eval_loss": 0.6882975697517395, "val_eval_accuracy":
0.5914489311163895, "val_eval_precision": 0.5510204081632653, "val_eval_recall":
0.5625, "val_eval_f1": 0.5567010309278351, "val_eval_runtime": 6.8434,
"val_eval_samples_per_second": 61.519, "val_eval_steps_per_second": 0.585,
"val_epoch": 2.0, "test_eval_loss": 0.6892734169960022, "test_eval_accuracy":
0.5430752453653217, "test_eval_precision": 0.5255813953488372,
"test_eval_recall": 0.5124716553287982, "test_eval_f1": 0.5189437428243399,
"test_eval_runtime": 9.3781, "test_eval_samples_per_second": 97.781,
"test_eval_steps_per_second": 0.853, "test_epoch": 2.0}}
{"timestamp": "2025-04-10T15:38:52.505714-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 2,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":

```

```

1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6880219578742981,
"train_eval_accuracy": 0.5477682067345341, "train_eval_precision":
0.5659777424483307, "train_eval_recall": 0.37503292072688965, "train_eval_f1":
0.451132583557738, "train_eval_runtime": 39.3665,
"train_eval_samples_per_second": 194.632, "train_eval_steps_per_second": 1.524,
"train_epoch": 2.0, "val_eval_loss": 0.6854250431060791, "val_eval_accuracy":
0.5795724465558195, "val_eval_precision": 0.5581395348837209, "val_eval_recall":
0.375, "val_eval_f1": 0.4485981308411215, "val_eval_runtime": 6.772,
"val_eval_samples_per_second": 62.168, "val_eval_steps_per_second": 0.591,
"val_epoch": 2.0, "test_eval_loss": 0.6856164932250977, "test_eval_accuracy":
0.5648854961832062, "test_eval_precision": 0.5729166666666666,
"test_eval_recall": 0.3741496598639456, "test_eval_f1": 0.45267489711934156,
"test_eval_runtime": 8.9617, "test_eval_samples_per_second": 102.324,
"test_eval_steps_per_second": 0.893, "test_epoch": 2.0}}
{"timestamp": "2025-04-10T15:46:37.337050-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,

```

```

"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.7801749110221863, "train_eval_accuracy":
0.5208822761681023, "train_eval_precision": 0.5156017830609212,
"train_eval_recall": 0.5483276270740058, "train_eval_f1": 0.5314613911933631,
"train_eval_runtime": 20.5254, "train_eval_samples_per_second": 373.294,
"train_eval_steps_per_second": 2.923, "train_epoch": 1.0, "val_eval_loss":
0.8155386447906494, "val_eval_accuracy": 0.501187648456057,
"val_eval_precision": 0.4605263157894737, "val_eval_recall": 0.546875,
"val_eval_f1": 0.5, "val_eval_runtime": 6.0876, "val_eval_samples_per_second":
69.157, "val_eval_steps_per_second": 0.657, "val_epoch": 1.0, "test_eval_loss":
0.7997273206710815, "test_eval_accuracy": 0.5103598691384951,
"test_eval_precision": 0.49166666666666664, "test_eval_recall":
0.5351473922902494, "test_eval_f1": 0.512486427795874, "test_eval_runtime":
7.2452, "test_eval_samples_per_second": 126.567, "test_eval_steps_per_second":
1.104, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T15:49:21.472385-07:00", "experiment_meta":

```

```

{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
5, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.6999151706695557, "train_eval_accuracy":
0.5370660401983817, "train_eval_precision": 0.5344162995594713,

```

```

"train_eval_recall": 0.5111930471424809, "train_eval_f1": 0.52254677614753,
"train_eval_runtime": 20.4485, "train_eval_samples_per_second": 374.697,
"train_eval_steps_per_second": 2.934, "train_epoch": 5.0, "val_eval_loss":
0.6929729580879211, "val_eval_accuracy": 0.5629453681710214,
"val_eval_precision": 0.5196078431372549, "val_eval_recall": 0.5520833333333334,
"val_eval_f1": 0.5353535353535354, "val_eval_runtime": 5.6174,
"val_eval_samples_per_second": 74.946, "val_eval_steps_per_second": 0.712,
"val_epoch": 5.0, "test_eval_loss": 0.7142435312271118, "test_eval_accuracy":
0.514721919302072, "test_eval_precision": 0.4953051643192488,
"test_eval_recall": 0.47845804988662133, "test_eval_f1": 0.4867358708189158,
"test_eval_runtime": 7.015, "test_eval_samples_per_second": 130.72,
"test_eval_steps_per_second": 1.14, "test_epoch": 5.0}}
{"timestamp": "2025-04-10T15:56:08.050290-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
15, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,

```

```

"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.6722455024719238, "train_eval_accuracy":
0.5891412163925868, "train_eval_precision": 0.5872546383436408,
"train_eval_recall": 0.57519094021596, "train_eval_f1": 0.581160191591272,
"train_eval_runtime": 20.5149, "train_eval_samples_per_second": 373.485,
"train_eval_steps_per_second": 2.925, "train_epoch": 15.0, "val_eval_loss":
0.6757411360740662, "val_eval_accuracy": 0.5724465558194775,
"val_eval_precision": 0.5309278350515464, "val_eval_recall": 0.5364583333333334,
"val_eval_f1": 0.533678756476684, "val_eval_runtime": 5.6554,
"val_eval_samples_per_second": 74.442, "val_eval_steps_per_second": 0.707,
"val_epoch": 15.0, "test_eval_loss": 0.679582953453064, "test_eval_accuracy":
0.5637949836423118, "test_eval_precision": 0.5475638051044084,
"test_eval_recall": 0.5351473922902494, "test_eval_f1": 0.5412844036697247,
"test_eval_runtime": 6.7341, "test_eval_samples_per_second": 136.173,
"test_eval_steps_per_second": 1.188, "test_epoch": 15.0}}
{"timestamp": "2025-04-10T15:57:21.078611-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
15, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,

```

```

"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.6722455024719238, "train_eval_accuracy":
0.5891412163925868, "train_eval_precision": 0.5872546383436408,
"train_eval_recall": 0.57519094021596, "train_eval_f1": 0.581160191591272,
"train_eval_runtime": 21.153, "train_eval_samples_per_second": 362.219,
"train_eval_steps_per_second": 2.836, "train_epoch": 15.0, "val_eval_loss":
0.6757411360740662, "val_eval_accuracy": 0.5724465558194775,
"val_eval_precision": 0.5309278350515464, "val_eval_recall": 0.5364583333333334,
"val_eval_f1": 0.533678756476684, "val_eval_runtime": 5.7126,
"val_eval_samples_per_second": 73.696, "val_eval_steps_per_second": 0.7,
"val_epoch": 15.0, "test_eval_loss": 0.679582953453064, "test_eval_accuracy":
0.5637949836423118, "test_eval_precision": 0.5475638051044084,
"test_eval_recall": 0.5351473922902494, "test_eval_f1": 0.5412844036697247,
"test_eval_runtime": 6.5727, "test_eval_samples_per_second": 139.517,
"test_eval_steps_per_second": 1.217, "test_epoch": 15.0}}
{"timestamp": "2025-04-10T16:01:36.764679-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-large", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.27.attn_norm.weight", "model.layers.27.attn.Wqkv.weight",
"model.layers.27.attn.Wo.weight", "model.layers.27.mlp_norm.weight",
"model.layers.27.mlp.Wi.weight", "model.layers.27.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 395833346, "model_trainable_params": 13309954,

```

```

"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-large",
"_commit_hash": "45bb4654a4d5aaff24dd11d4781fa46d39bf8c13",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 1024, "intermediate_size": 2624,
"num_hidden_layers": 28, "num_attention_heads": 16, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.7490190267562866, "train_eval_accuracy":
0.5016966849386583, "train_eval_precision": 0.49730838246603437,
"train_eval_recall": 0.5109296813273637, "train_eval_f1": 0.5040270200051962,
"train_eval_runtime": 46.1965, "train_eval_samples_per_second": 165.857,
"train_eval_steps_per_second": 1.299, "train_epoch": 1.0, "val_eval_loss":
0.7550219893455505, "val_eval_accuracy": 0.5154394299287411,
"val_eval_precision": 0.47345132743362833, "val_eval_recall":
0.5572916666666666, "val_eval_f1": 0.5119617224880383, "val_eval_runtime":
7.1355, "val_eval_samples_per_second": 59.0, "val_eval_steps_per_second": 0.561,
"val_epoch": 1.0, "test_eval_loss": 0.7513988614082336, "test_eval_accuracy":
0.5125408942202835, "test_eval_precision": 0.49416342412451364,

```



```

"test_eval_recall": 0.5759637188208617, "test_eval_f1": 0.5319371727748691,
"test_eval_runtime": 9.7168, "test_eval_samples_per_second": 94.373,
"test_eval_steps_per_second": 0.823, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T16:13:30.122949-07:00", "experiment_meta":
{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",
"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",
"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":

```

```

"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":
null, "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":
"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":
12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,
"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.698921263217926, "train_eval_accuracy":
0.4986948577394936, "train_eval_precision": 0.49661121380160195,
"train_eval_recall": 0.8490913879378457, "train_eval_f1": 0.6266886966663427,
"train_eval_runtime": 19.5519, "train_eval_samples_per_second": 391.88,
"train_eval_steps_per_second": 3.069, "train_epoch": 1.0, "val_eval_loss":
0.7018232345581055, "val_eval_accuracy": 0.48931116389548696,
"val_eval_precision": 0.46831955922865015, "val_eval_recall":
0.8854166666666666, "val_eval_f1": 0.6126126126126126, "val_eval_runtime":
5.5759, "val_eval_samples_per_second": 75.504, "val_eval_steps_per_second":
0.717, "val_epoch": 1.0, "test_eval_loss": 0.701395571231842,
"test_eval_accuracy": 0.48091603053435117, "test_eval_precision":
0.47736093143596375, "test_eval_recall": 0.8367346938775511, "test_eval_f1":
0.6079077429983526, "test_eval_runtime": 6.4673, "test_eval_samples_per_second":
141.79, "test_eval_steps_per_second": 1.237, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T16:16:24.806773-07:00", "experiment_meta":
{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 3,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",
"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",
"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":

```

```

{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":
"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":
null, "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":
"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":
12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,
"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.6927720904350281, "train_eval_accuracy":
0.5053510832680762, "train_eval_precision": 0.5309734513274337,
"train_eval_recall": 0.015801948907031866, "train_eval_f1":
0.030690537084398978, "train_eval_runtime": 19.4037,
"train_eval_samples_per_second": 394.874, "train_eval_steps_per_second": 3.092,
"train_epoch": 3.0, "val_eval_loss": 0.6913222074508667, "val_eval_accuracy":
0.5534441805225653, "val_eval_precision": 1.0, "val_eval_recall":
0.020833333333333332, "val_eval_f1": 0.04081632653061224, "val_eval_runtime":
5.7043, "val_eval_samples_per_second": 73.804, "val_eval_steps_per_second":
0.701, "val_epoch": 3.0, "test_eval_loss": 0.6919171810150146,
"test_eval_accuracy": 0.5288985823336968, "test_eval_precision":
0.7142857142857143, "test_eval_recall": 0.034013605442176874, "test_eval_f1":
0.06493506493506493, "test_eval_runtime": 6.5459,
"test_eval_samples_per_second": 140.087, "test_eval_steps_per_second": 1.222,
"test_epoch": 3.0}}

```

```

{"timestamp": "2025-04-10T16:20:29.246186-07:00", "experiment_meta":
{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 5,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",
"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",
"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":
"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":
null, "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":

```

```

"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":
12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,
"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.692355215549469, "train_eval_accuracy":
0.5116157661185069, "train_eval_precision": 0.6341463414634146,
"train_eval_recall": 0.034237555965235714, "train_eval_f1": 0.06496751624187906,
"train_eval_runtime": 19.4226, "train_eval_samples_per_second": 394.49,
"train_eval_steps_per_second": 3.089, "train_epoch": 5.0, "val_eval_loss":
0.6912702322006226, "val_eval_accuracy": 0.5439429928741093,
"val_eval_precision": 0.5, "val_eval_recall": 0.015625, "val_eval_f1":
0.030303030303030304, "val_eval_runtime": 5.5579, "val_eval_samples_per_second":
75.748, "val_eval_steps_per_second": 0.72, "val_epoch": 5.0, "test_eval_loss":
0.6918637752532959, "test_eval_accuracy": 0.52453653217012,
"test_eval_precision": 0.6, "test_eval_recall": 0.034013605442176874,
"test_eval_f1": 0.06437768240343347, "test_eval_runtime": 6.5673,
"test_eval_samples_per_second": 139.631, "test_eval_steps_per_second": 1.218,
"test_epoch": 5.0}}
{"timestamp": "2025-04-10T16:33:56.812453-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.11.rel_attn.q", "transformer.layer.11.rel_attn.k",
"transformer.layer.11.rel_attn.v", "transformer.layer.11.rel_attn.o",
"transformer.layer.11.rel_attn.r", "transformer.layer.11.rel_attn.r_r_bias",
"transformer.layer.11.rel_attn.r_s_bias",
"transformer.layer.11.rel_attn.r_w_bias",
"transformer.layer.11.rel_attn.seg_embed",
"transformer.layer.11.rel_attn.layer_norm.weight",
"transformer.layer.11.rel_attn.layer_norm.bias",
"transformer.layer.11.ff.layer_norm.weight",
"transformer.layer.11.ff.layer_norm.bias",
"transformer.layer.11.ff.layer_1.weight",
"transformer.layer.11.ff.layer_1.bias",
"transformer.layer.11.ff.layer_2.weight",
"transformer.layer.11.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 117310466, "model_trainable_params":
8270594, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 768, "n_layer": 12, "n_head": 12, "d_head": 64,
"ff_activation": "gelu", "d_inner": 3072, "untie_r": true, "attn_type": "bi",
"initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1, "mem_len":
null, "reuse_len": null, "bi_data": false, "clamp_len": -1, "same_length":
false, "summary_type": "last", "summary_use_proj": true, "summary_activation":

```

```

"tanh", "summary_last_dropout": 0.1, "start_n_top": 5, "end_n_top": 5,
"bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2, "use_mems_eval": true,
"use_mems_train": false, "return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-base-cased",
"_commit_hash": "ceaa69c7bc5e512b5007106a7ccbb7daf24b2c79",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}, {"run_metrics":
{"train_eval_loss": 0.7004867196083069, "train_eval_accuracy":
0.486948577394936, "train_eval_precision": 0.48018923713778827,
"train_eval_recall": 0.4277060837503292, "train_eval_f1": 0.45243070065468727,
"train_eval_runtime": 19.8022, "train_eval_samples_per_second": 386.927,
"train_eval_steps_per_second": 3.03, "train_epoch": 1.0, "val_eval_loss":
0.6955275535583496, "val_eval_accuracy": 0.501187648456057,
"val_eval_precision": 0.449438202247191, "val_eval_recall": 0.4166666666666667,
"val_eval_f1": 0.43243243243243246, "val_eval_runtime": 2.3815,
"val_eval_samples_per_second": 176.779, "val_eval_steps_per_second": 1.68,
"val_epoch": 1.0, "test_eval_loss": 0.7008840441703796, "test_eval_accuracy":
0.48091603053435117, "test_eval_precision": 0.458628841607565,
"test_eval_recall": 0.4399092970521542, "test_eval_f1": 0.44907407407407407,
"test_eval_runtime": 3.5375, "test_eval_samples_per_second": 259.22,
"test_eval_steps_per_second": 2.261, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T16:42:28.022200-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.23.rel_attn.q", "transformer.layer.23.rel_attn.k",
"transformer.layer.23.rel_attn.v", "transformer.layer.23.rel_attn.o",
"transformer.layer.23.rel_attn.r", "transformer.layer.23.rel_attn.r_r_bias",
"transformer.layer.23.rel_attn.r_s_bias",

```

```

"transformer.layer.23.rel_attn.r_w_bias",
"transformer.layer.23.rel_attn.seg_embed",
"transformer.layer.23.rel_attn.layer_norm.weight",
"transformer.layer.23.rel_attn.layer_norm.bias",
"transformer.layer.23.ff.layer_norm.weight",
"transformer.layer.23.ff.layer_norm.bias",
"transformer.layer.23.ff.layer_1.weight",
"transformer.layer.23.ff.layer_1.bias",
"transformer.layer.23.ff.layer_2.weight",
"transformer.layer.23.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 361320450, "model_trainable_params":
14697474, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 1024, "n_layer": 24, "n_head": 16, "d_head":
64, "ff_activation": "gelu", "d_inner": 4096, "untie_r": true, "attn_type":
"bi", "initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1,
"mem_len": null, "reuse_len": null, "bi_data": false, "clamp_len": -1,
"same_length": false, "summary_type": "last", "summary_use_proj": true,
"summary_activation": "tanh", "summary_last_dropout": 0.1, "start_n_top": 5,
"end_n_top": 5, "bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2,
"use_mems_eval": true, "use_mems_train": false, "return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-large-cased",
"_commit_hash": "37658b4f179eaf971d127d16bfb9ca676a93034",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}, "run_metrics":
{"train_eval_loss": 0.7052111029624939, "train_eval_accuracy":
0.5082223962411903, "train_eval_precision": 0.5036923860453272,
"train_eval_recall": 0.5209375823018172, "train_eval_f1": 0.5121698601760746,
"train_eval_runtime": 54.1724, "train_eval_samples_per_second": 141.437,

```

```

"train_eval_steps_per_second": 1.108, "train_epoch": 1.0, "val_eval_loss":
0.7174055576324463, "val_eval_accuracy": 0.4821852731591449,
"val_eval_precision": 0.4375, "val_eval_recall": 0.4739583333333333,
"val_eval_f1": 0.455, "val_eval_runtime": 4.2671, "val_eval_samples_per_second":
98.662, "val_eval_steps_per_second": 0.937, "val_epoch": 1.0, "test_eval_loss":
0.706987738609314, "test_eval_accuracy": 0.4961832061068702,
"test_eval_precision": 0.47761194029850745, "test_eval_recall":
0.5079365079365079, "test_eval_f1": 0.49230769230769234, "test_eval_runtime":
7.7045, "test_eval_samples_per_second": 119.021, "test_eval_steps_per_second":
1.038, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T17:22:38.104957-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7055578827857971, "train_eval_accuracy": 0.4871456822676335,
"train_eval_precision": 0.4931693989071038, "train_eval_recall":

```



```

0.9525065963060686, "train_eval_f1": 0.6498649864986499, "train_eval_runtime":
6.9876, "train_eval_samples_per_second": 217.098, "train_eval_steps_per_second":
1.717, "train_epoch": 1.0, "val_eval_loss": 0.6994861364364624,
"val_eval_accuracy": 0.5252525252525253, "val_eval_precision":
0.5204081632653061, "val_eval_recall": 1.0, "val_eval_f1": 0.6845637583892618,
"val_eval_runtime": 5.3701, "val_eval_samples_per_second": 18.435,
"val_eval_steps_per_second": 0.186, "val_epoch": 1.0, "test_eval_loss":
0.6954776644706726, "test_eval_accuracy": 0.532608695652174,
"test_eval_precision": 0.5359116022099447, "test_eval_recall":
0.9797979797979798, "test_eval_f1": 0.6928571428571428, "test_eval_runtime":
5.1964, "test_eval_samples_per_second": 35.409, "test_eval_steps_per_second":
0.385, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T17:23:30.216271-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,

```

```

"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6943163275718689,
"train_eval_accuracy": 0.4996704021094265, "train_eval_precision":
0.4996704021094265, "train_eval_recall": 1.0, "train_eval_f1":
0.6663736263736264, "train_eval_runtime": 7.0753,
"train_eval_samples_per_second": 214.408, "train_eval_steps_per_second": 1.696,
"train_epoch": 1.0, "val_eval_loss": 0.6930510401725769, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 8.9879,
"val_eval_samples_per_second": 11.015, "val_eval_steps_per_second": 0.111,
"val_epoch": 1.0, "test_eval_loss": 0.6911553740501404, "test_eval_accuracy":
0.5380434782608695, "test_eval_precision": 0.5380434782608695,
"test_eval_recall": 1.0, "test_eval_f1": 0.6996466431095406,
"test_eval_runtime": 4.9592, "test_eval_samples_per_second": 37.103,
"test_eval_steps_per_second": 0.403, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T17:27:20.459219-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,

```

```

"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaeff65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.7026900053024292,
"train_eval_accuracy": 0.5187870797626896, "train_eval_precision": 0.57,
"train_eval_recall": 0.1503957783641161, "train_eval_f1": 0.23799582463465555,
"train_eval_runtime": 12.0912, "train_eval_samples_per_second": 125.463,
"train_eval_steps_per_second": 0.992, "train_epoch": 1.0, "val_eval_loss":
0.7341456413269043, "val_eval_accuracy": 0.45454545454545453,
"val_eval_precision": 0.4, "val_eval_recall": 0.11764705882352941,
"val_eval_f1": 0.18181818181818182, "val_eval_runtime": 5.6239,
"val_eval_samples_per_second": 17.603, "val_eval_steps_per_second": 0.178,
"val_epoch": 1.0, "test_eval_loss": 0.7290977835655212, "test_eval_accuracy":
0.4891304347826087, "test_eval_precision": 0.6, "test_eval_recall":
0.15151515151515152, "test_eval_f1": 0.24193548387096775, "test_eval_runtime":
5.8453, "test_eval_samples_per_second": 31.478, "test_eval_steps_per_second":
0.342, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T17:28:31.395352-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",

```

```

"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6972289085388184,
"train_eval_accuracy": 0.4996704021094265, "train_eval_precision":
0.4996704021094265, "train_eval_recall": 1.0, "train_eval_f1":
0.6663736263736264, "train_eval_runtime": 11.896,
"train_eval_samples_per_second": 127.522, "train_eval_steps_per_second": 1.009,
"train_epoch": 1.0, "val_eval_loss": 0.6957302093505859, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 6.4878,
"val_eval_samples_per_second": 15.259, "val_eval_steps_per_second": 0.154,
"val_epoch": 1.0, "test_eval_loss": 0.6913740038871765, "test_eval_accuracy":

```

```

0.5380434782608695, "test_eval_precision": 0.5380434782608695,
"test_eval_recall": 1.0, "test_eval_f1": 0.6996466431095406,
"test_eval_runtime": 6.0399, "test_eval_samples_per_second": 30.464,
"test_eval_steps_per_second": 0.331, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T17:29:38.218189-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":

```

```

false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 2.545469284057617, "train_eval_accuracy":
0.4976928147659855, "train_eval_precision": 0.4986789960369881,
"train_eval_recall": 0.996042216358839, "train_eval_f1": 0.664612676056338,
"train_eval_runtime": 8.2857, "train_eval_samples_per_second": 183.087,
"train_eval_steps_per_second": 1.448, "train_epoch": 1.0, "val_eval_loss":
2.659238815307617, "val_eval_accuracy": 0.5151515151515151,
"val_eval_precision": 0.5151515151515151, "val_eval_recall": 1.0, "val_eval_f1":
0.68, "val_eval_runtime": 5.1564, "val_eval_samples_per_second": 19.2,
"val_eval_steps_per_second": 0.194, "val_epoch": 1.0, "test_eval_loss":
2.2664828300476074, "test_eval_accuracy": 0.5434782608695652,
"test_eval_precision": 0.5409836065573771, "test_eval_recall": 1.0,
"test_eval_f1": 0.7021276595744681, "test_eval_runtime": 5.6613,
"test_eval_samples_per_second": 32.501, "test_eval_steps_per_second": 0.353,
"test_epoch": 1.0}}
{"timestamp": "2025-04-10T17:30:52.763212-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-large", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.27.attn_norm.weight", "model.layers.27.attn.Wqkv.weight",
"model.layers.27.attn.Wo.weight", "model.layers.27.mlp_norm.weight",
"model.layers.27.mlp.Wi.weight", "model.layers.27.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 395833346, "model_trainable_params": 13309954,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-large",

```

```

"_commit_hash": "45bb4654a4d5aaff24dd11d4781fa46d39bf8c13",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 1024, "intermediate_size": 2624,
"num_hidden_layers": 28, "num_attention_heads": 16, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.8578765392303467, "train_eval_accuracy":
0.5042847725774555, "train_eval_precision": 0.5020188425302826,
"train_eval_recall": 0.9841688654353562, "train_eval_f1": 0.6648841354723708,
"train_eval_runtime": 13.1193, "train_eval_samples_per_second": 115.632,
"train_eval_steps_per_second": 0.915, "train_epoch": 1.0, "val_eval_loss":
0.8566176891326904, "val_eval_accuracy": 0.5353535353535354,
"val_eval_precision": 0.5257731958762887, "val_eval_recall": 1.0, "val_eval_f1":
0.6891891891891891, "val_eval_runtime": 5.9247, "val_eval_samples_per_second":
16.71, "val_eval_steps_per_second": 0.169, "val_epoch": 1.0, "test_eval_loss":
0.8890158534049988, "test_eval_accuracy": 0.5380434782608695,
"test_eval_precision": 0.5384615384615384, "test_eval_recall": 0.98989898989899,
"test_eval_f1": 0.697508896797153, "test_eval_runtime": 6.204,
"test_eval_samples_per_second": 29.658, "test_eval_steps_per_second": 0.322,
"test_epoch": 1.0}}
{"timestamp": "2025-04-10T17:31:46.893363-07:00", "experiment_meta":
{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",
"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",

```

```

"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":
"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":
null, "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":
"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":
12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,
"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.7042434215545654, "train_eval_accuracy":
0.5016479894528675, "train_eval_precision": 0.5125, "train_eval_recall":
0.05408970976253298, "train_eval_f1": 0.09785202863961814, "train_eval_runtime":
9.1051, "train_eval_samples_per_second": 166.61, "train_eval_steps_per_second":
1.318, "train_epoch": 1.0, "val_eval_loss": 0.7114705443382263,
"val_eval_accuracy": 0.5050505050505051, "val_eval_precision":
0.6666666666666666, "val_eval_recall": 0.0784313725490196, "val_eval_f1":
0.14035087719298245, "val_eval_runtime": 5.2922, "val_eval_samples_per_second":
18.707, "val_eval_steps_per_second": 0.189, "val_epoch": 1.0, "test_eval_loss":
0.7180479168891907, "test_eval_accuracy": 0.46195652173913043,

```



```

"test_eval_precision": 0.5, "test_eval_recall": 0.04040404040404041,
"test_eval_f1": 0.07476635514018691, "test_eval_runtime": 5.1994,
"test_eval_samples_per_second": 35.389, "test_eval_steps_per_second": 0.385,
"test_epoch": 1.0}}
{"timestamp": "2025-04-10T17:35:58.847095-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.11.rel_attn.q", "transformer.layer.11.rel_attn.k",
"transformer.layer.11.rel_attn.v", "transformer.layer.11.rel_attn.o",
"transformer.layer.11.rel_attn.r", "transformer.layer.11.rel_attn.r_r_bias",
"transformer.layer.11.rel_attn.r_s_bias",
"transformer.layer.11.rel_attn.r_w_bias",
"transformer.layer.11.rel_attn.seg_embed",
"transformer.layer.11.rel_attn.layer_norm.weight",
"transformer.layer.11.rel_attn.layer_norm.bias",
"transformer.layer.11.ff.layer_norm.weight",
"transformer.layer.11.ff.layer_norm.bias",
"transformer.layer.11.ff.layer_1.weight",
"transformer.layer.11.ff.layer_1.bias",
"transformer.layer.11.ff.layer_2.weight",
"transformer.layer.11.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 117310466, "model_trainable_params":
8270594, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 768, "n_layer": 12, "n_head": 12, "d_head": 64,
"ff_activation": "gelu", "d_inner": 3072, "untie_r": true, "attn_type": "bi",
"initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1, "mem_len":
null, "reuse_len": null, "bi_data": false, "clamp_len": -1, "same_length":
false, "summary_type": "last", "summary_use_proj": true, "summary_activation":
"tanh", "summary_last_dropout": 0.1, "start_n_top": 5, "end_n_top": 5,
"bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2, "use_mems_eval": true,
"use_mems_train": false, "return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":

```

```

"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-base-cased",
"_commit_hash": "ceaa69c7bc5e512b5007106a7ccbb7daf24b2c79",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}, "run_metrics":
{"train_eval_loss": 0.8484305739402771, "train_eval_accuracy":
0.5023071852340145, "train_eval_precision": 1.0, "train_eval_recall":
0.00395778364116095, "train_eval_f1": 0.00788436268068331, "train_eval_runtime":
8.9289, "train_eval_samples_per_second": 169.898, "train_eval_steps_per_second":
1.344, "train_epoch": 1.0, "val_eval_loss": 0.8675015568733215,
"val_eval_accuracy": 0.48484848484848486, "val_eval_precision": 0.0,
"val_eval_recall": 0.0, "val_eval_f1": 0.0, "val_eval_runtime": 5.4925,
"val_eval_samples_per_second": 18.024, "val_eval_steps_per_second": 0.182,
"val_epoch": 1.0, "test_eval_loss": 0.8848596811294556, "test_eval_accuracy":
0.46195652173913043, "test_eval_precision": 0.0, "test_eval_recall": 0.0,
"test_eval_f1": 0.0, "test_eval_runtime": 5.6157,
"test_eval_samples_per_second": 32.765, "test_eval_steps_per_second": 0.356,
"test_epoch": 1.0}}
{"timestamp": "2025-04-10T17:37:20.011618-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.23.rel_attn.q", "transformer.layer.23.rel_attn.k",
"transformer.layer.23.rel_attn.v", "transformer.layer.23.rel_attn.o",
"transformer.layer.23.rel_attn.r", "transformer.layer.23.rel_attn.r_r_bias",
"transformer.layer.23.rel_attn.r_s_bias",
"transformer.layer.23.rel_attn.r_w_bias",
"transformer.layer.23.rel_attn.seg_embed",
"transformer.layer.23.rel_attn.layer_norm.weight",
"transformer.layer.23.rel_attn.layer_norm.bias",
"transformer.layer.23.ff.layer_norm.weight",
"transformer.layer.23.ff.layer_norm.bias",
"transformer.layer.23.ff.layer_1.weight",
"transformer.layer.23.ff.layer_1.bias",
"transformer.layer.23.ff.layer_2.weight",
"transformer.layer.23.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 361320450, "model_trainable_params":
14697474, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 1024, "n_layer": 24, "n_head": 16, "d_head":
64, "ff_activation": "gelu", "d_inner": 4096, "untie_r": true, "attn_type":
"bi", "initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1,
"mem_len": null, "reuse_len": null, "bi_data": false, "clamp_len": -1,
"same_length": false, "summary_type": "last", "summary_use_proj": true,
"summary_activation": "tanh", "summary_last_dropout": 0.1, "start_n_top": 5,

```

```

"end_n_top": 5, "bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2,
"use_mems_eval": true, "use_mems_train": false, "return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-large-cased",
"_commit_hash": "37658b4f179eaf971d127d16bfbd9ca676a93034",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}, "run_metrics":
{"train_eval_loss": 0.6847349405288696, "train_eval_accuracy":
0.5477916941331575, "train_eval_precision": 0.5454545454545454,
"train_eval_recall": 0.5699208443271768, "train_eval_f1": 0.5574193548387096,
"train_eval_runtime": 15.6086, "train_eval_samples_per_second": 97.19,
"train_eval_steps_per_second": 0.769, "train_epoch": 1.0, "val_eval_loss":
0.6953040361404419, "val_eval_accuracy": 0.494949494949495,
"val_eval_precision": 0.5079365079365079, "val_eval_recall": 0.6274509803921569,
"val_eval_f1": 0.5614035087719298, "val_eval_runtime": 6.5167,
"val_eval_samples_per_second": 15.192, "val_eval_steps_per_second": 0.153,
"val_epoch": 1.0, "test_eval_loss": 0.7078865766525269, "test_eval_accuracy":
0.4945652173913043, "test_eval_precision": 0.5294117647058824,
"test_eval_recall": 0.5454545454545454, "test_eval_f1": 0.5373134328358209,
"test_eval_runtime": 6.451, "test_eval_samples_per_second": 28.523,
"test_eval_steps_per_second": 0.31, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T17:55:21.671457-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity_75th_split",
"layers_to_unfreeze": ["bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 7680002, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",

```

```

"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.5472237467765808, "train_eval_accuracy": 0.7633777081701906,
"train_eval_precision": 0.0, "train_eval_recall": 0.0, "train_eval_f1": 0.0,
"train_eval_runtime": 16.7137, "train_eval_samples_per_second": 458.427,
"train_eval_steps_per_second": 3.59, "train_epoch": 1.0, "val_eval_loss":
0.5287129878997803, "val_eval_accuracy": 0.7790973871733967,
"val_eval_precision": 0.0, "val_eval_recall": 0.0, "val_eval_f1": 0.0,
"val_eval_runtime": 5.5052, "val_eval_samples_per_second": 76.472,
"val_eval_steps_per_second": 0.727, "val_epoch": 1.0, "test_eval_loss":
0.505957841873169, "test_eval_accuracy": 0.7960741548527808,
"test_eval_precision": 0.0, "test_eval_recall": 0.0, "test_eval_f1": 0.0,
"test_eval_runtime": 6.7653, "test_eval_samples_per_second": 135.545,
"test_eval_steps_per_second": 1.183, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T17:56:33.142167-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity_75th_split",
"layers_to_unfreeze": ["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",

```

```

"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6546781659126282,
"train_eval_accuracy": 0.7633777081701906, "train_eval_precision": 0.0,
"train_eval_recall": 0.0, "train_eval_f1": 0.0, "train_eval_runtime": 16.0424,
"train_eval_samples_per_second": 477.61, "train_eval_steps_per_second": 3.74,

```

```

"train_epoch": 1.0, "val_eval_loss": 0.6522068977355957, "val_eval_accuracy":
0.7790973871733967, "val_eval_precision": 0.0, "val_eval_recall": 0.0,
"val_eval_f1": 0.0, "val_eval_runtime": 5.3528, "val_eval_samples_per_second":
78.651, "val_eval_steps_per_second": 0.747, "val_epoch": 1.0, "test_eval_loss":
0.6492957472801208, "test_eval_accuracy": 0.7960741548527808,
"test_eval_precision": 0.0, "test_eval_recall": 0.0, "test_eval_f1": 0.0,
"test_eval_runtime": 6.4138, "test_eval_samples_per_second": 142.973,
"test_eval_steps_per_second": 1.247, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T18:04:02.839207-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity_75th_split",
"layers_to_unfreeze": ["bert.encoder.layer.23.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 333581314,
"model_trainable_params": 13647874, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaef65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.5770174860954285,

```

```

"train_eval_accuracy": 0.7633777081701906, "train_eval_precision": 0.5,
"train_eval_recall": 0.0005515719801434088, "train_eval_f1":
0.0011019283746556473, "train_eval_runtime": 39.2441,
"train_eval_samples_per_second": 195.24, "train_eval_steps_per_second": 1.529,
"train_epoch": 1.0, "val_eval_loss": 0.565083920955658, "val_eval_accuracy":
0.7790973871733967, "val_eval_precision": 0.0, "val_eval_recall": 0.0,
"val_eval_f1": 0.0, "val_eval_runtime": 6.787, "val_eval_samples_per_second":
62.031, "val_eval_steps_per_second": 0.589, "val_epoch": 1.0, "test_eval_loss":
0.5492461323738098, "test_eval_accuracy": 0.7960741548527808,
"test_eval_precision": 0.0, "test_eval_recall": 0.0, "test_eval_f1": 0.0,
"test_eval_runtime": 9.329, "test_eval_samples_per_second": 98.295,
"test_eval_steps_per_second": 0.858, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T18:06:20.082953-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity_75th_split",
"layers_to_unfreeze": ["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,

```

```

"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6395546197891235,
"train_eval_accuracy": 0.7633777081701906, "train_eval_precision": 0.0,
"train_eval_recall": 0.0, "train_eval_f1": 0.0, "train_eval_runtime": 40.126,
"train_eval_samples_per_second": 190.949, "train_eval_steps_per_second": 1.495,
"train_epoch": 1.0, "val_eval_loss": 0.6359410881996155, "val_eval_accuracy":
0.7790973871733967, "val_eval_precision": 0.0, "val_eval_recall": 0.0,
"val_eval_f1": 0.0, "val_eval_runtime": 6.6799, "val_eval_samples_per_second":
63.025, "val_eval_steps_per_second": 0.599, "val_epoch": 1.0, "test_eval_loss":
0.6321974992752075, "test_eval_accuracy": 0.7960741548527808,
"test_eval_precision": 0.0, "test_eval_recall": 0.0, "test_eval_f1": 0.0,
"test_eval_runtime": 9.0385, "test_eval_samples_per_second": 101.455,
"test_eval_steps_per_second": 0.885, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T18:07:58.072600-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity_75th_split",
"layers_to_unfreeze": ["model.layers.21.attn_norm.weight",
"model.layers.21.attn.Wqkv.weight", "model.layers.21.attn.Wo.weight",
"model.layers.21.mlp_norm.weight", "model.layers.21.mlp.Wi.weight",
"model.layers.21.mlp.Wo.weight", "model.final_norm.weight", "head.dense.weight",
"head.norm.weight", "classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,

```



```

"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 2.5128276348114014, "train_eval_accuracy":
0.25489428347689896, "train_eval_precision": 0.23859366612989802,
"train_eval_recall": 0.9806949806949807, "train_eval_f1": 0.3838100377765785,
"train_eval_runtime": 20.2079, "train_eval_samples_per_second": 379.158,
"train_eval_steps_per_second": 2.969, "train_epoch": 1.0, "val_eval_loss":
2.580798387527466, "val_eval_accuracy": 0.22565320665083136,
"val_eval_precision": 0.21654501216545013, "val_eval_recall": 0.956989247311828,
"val_eval_f1": 0.3531746031746032, "val_eval_runtime": 5.6225,
"val_eval_samples_per_second": 74.878, "val_eval_steps_per_second": 0.711,
"val_epoch": 1.0, "test_eval_loss": 2.600764274597168, "test_eval_accuracy":
0.22464558342420937, "test_eval_precision": 0.20495495495495494,
"test_eval_recall": 0.9732620320855615, "test_eval_f1": 0.3386046511627907,
"test_eval_runtime": 7.1035, "test_eval_samples_per_second": 129.091,
"test_eval_steps_per_second": 1.126, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T18:10:33.131556-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-large", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity_75th_split",
"layers_to_unfreeze": ["model.layers.27.attn_norm.weight",
"model.layers.27.attn.Wqkv.weight", "model.layers.27.attn.Wo.weight",
"model.layers.27.mlp_norm.weight", "model.layers.27.mlp.Wi.weight"],

```

```

"model.layers.27.mlp.Wo.weight", "model.final_norm.weight", "head.dense.weight",
"head.norm.weight", "classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 395833346, "model_trainable_params": 13309954,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-large",
"_commit_hash": "45bb4654a4d5aaff24dd11d4781fa46d39bf8c13",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 1024, "intermediate_size": 2624,
"num_hidden_layers": 28, "num_attention_heads": 16, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.7144438028335571, "train_eval_accuracy":
0.5203602192638997, "train_eval_precision": 0.23369565217391305,
"train_eval_recall": 0.4506343077771649, "train_eval_f1": 0.3077792427952533,
"train_eval_runtime": 47.1027, "train_eval_samples_per_second": 162.666,
"train_eval_steps_per_second": 1.274, "train_epoch": 1.0, "val_eval_loss":
0.6973762512207031, "val_eval_accuracy": 0.5415676959619953,
"val_eval_precision": 0.2222222222222222, "val_eval_recall":
0.43010752688172044, "val_eval_f1": 0.29304029304029305, "val_eval_runtime":

```

```

7.5972, "val_eval_samples_per_second": 55.415, "val_eval_steps_per_second":
0.527, "val_epoch": 1.0, "test_eval_loss": 0.7029761672019958,
"test_eval_accuracy": 0.5267175572519084, "test_eval_precision":
0.20803782505910165, "test_eval_recall": 0.47058823529411764, "test_eval_f1":
0.28852459016393445, "test_eval_runtime": 9.7497,
"test_eval_samples_per_second": 94.054, "test_eval_steps_per_second": 0.821,
"test_epoch": 1.0}}
{"timestamp": "2025-04-10T18:12:07.830546-07:00", "experiment_meta":
{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity_75th_split",
"layers_to_unfreeze":
["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",
"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",
"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":

```

```

{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":
"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":
null, "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":
"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":
12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,
"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.5560451745986938, "train_eval_accuracy":
0.7635082223962412, "train_eval_precision": 0.5384615384615384,
"train_eval_recall": 0.003861003861003861, "train_eval_f1":
0.007667031763417305, "train_eval_runtime": 19.5349,
"train_eval_samples_per_second": 392.22, "train_eval_steps_per_second": 3.071,
"train_epoch": 1.0, "val_eval_loss": 0.5420940518379211, "val_eval_accuracy":
0.7767220902612827, "val_eval_precision": 0.0, "val_eval_recall": 0.0,
"val_eval_f1": 0.0, "val_eval_runtime": 5.6098, "val_eval_samples_per_second":
75.047, "val_eval_steps_per_second": 0.713, "val_epoch": 1.0, "test_eval_loss":
0.5194469690322876, "test_eval_accuracy": 0.7960741548527808,
"test_eval_precision": 0.0, "test_eval_recall": 0.0, "test_eval_f1": 0.0,
"test_eval_runtime": 6.4845, "test_eval_samples_per_second": 141.414,
"test_eval_steps_per_second": 1.234, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T21:43:02.861600-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,

```

```

"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6933034062385559, "train_eval_accuracy": 0.5080918820151397,
"train_eval_precision": 0.5032558139534884, "train_eval_recall":
0.5699236239136161, "train_eval_f1": 0.5345189576386316, "train_eval_runtime":
12.1661, "train_eval_samples_per_second": 629.78, "train_eval_steps_per_second":
4.932, "train_epoch": 1.0, "val_eval_loss": 0.6924113035202026,
"val_eval_accuracy": 0.5296912114014252, "val_eval_precision":
0.48717948717948717, "val_eval_recall": 0.59375, "val_eval_f1":
0.5352112676056338, "val_eval_runtime": 1.6433, "val_eval_samples_per_second":
256.188, "val_eval_steps_per_second": 2.434, "val_epoch": 1.0, "test_eval_loss":
0.6913138628005981, "test_eval_accuracy": 0.5452562704471101,
"test_eval_precision": 0.5234375, "test_eval_recall": 0.6077097505668935,
"test_eval_f1": 0.5624344176285414, "test_eval_runtime": 2.3996,
"test_eval_samples_per_second": 382.149, "test_eval_steps_per_second": 3.334,
"test_epoch": 1.0}}
{"timestamp": "2025-04-10T21:43:45.835430-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",

```

```

"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6931643486022949,
"train_eval_accuracy": 0.49543200208822763, "train_eval_precision":
0.49546708711076076, "train_eval_recall": 0.9931524888069528, "train_eval_f1":
0.6611150070126227, "train_eval_runtime": 12.203,
"train_eval_samples_per_second": 627.88, "train_eval_steps_per_second": 4.917,
"train_epoch": 1.0, "val_eval_loss": 0.6936803460121155, "val_eval_accuracy":
0.45605700712589076, "val_eval_precision": 0.45605700712589076,
"val_eval_recall": 1.0, "val_eval_f1": 0.6264274061990212, "val_eval_runtime":
1.9975, "val_eval_samples_per_second": 210.759, "val_eval_steps_per_second":
2.002, "val_epoch": 1.0, "test_eval_loss": 0.693478524684906,
"test_eval_accuracy": 0.4776444929116685, "test_eval_precision":
0.47912087912087914, "test_eval_recall": 0.9886621315192744, "test_eval_f1":
0.6454478164322723, "test_eval_runtime": 2.5075, "test_eval_samples_per_second":

```

```

365.701, "test_eval_steps_per_second": 3.19, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T21:47:05.320003-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaeff65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.7082446217536926,
"train_eval_accuracy": 0.5071782824327852, "train_eval_precision":
0.5549738219895288, "train_eval_recall": 0.027916776402422964, "train_eval_f1":
0.05315947843530592, "train_eval_runtime": 35.8453,
"train_eval_samples_per_second": 213.752, "train_eval_steps_per_second": 1.674,
"train_epoch": 1.0, "val_eval_loss": 0.694320559501648, "val_eval_accuracy":
0.5510688836104513, "val_eval_precision": 0.7142857142857143, "val_eval_recall":
0.026041666666666668, "val_eval_f1": 0.05025125628140704, "val_eval_runtime":
3.4796, "val_eval_samples_per_second": 120.992, "val_eval_steps_per_second":

```

```

1.15, "val_epoch": 1.0, "test_eval_loss": 0.7033457159996033,
"test_eval_accuracy": 0.5223555070883316, "test_eval_precision":
0.5652173913043478, "test_eval_recall": 0.02947845804988662, "test_eval_f1":
0.05603448275862069, "test_eval_runtime": 5.4953,
"test_eval_samples_per_second": 166.869, "test_eval_steps_per_second": 1.456,
"test_epoch": 1.0}}
{"timestamp": "2025-04-10T21:48:54.281828-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,

```



```

"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdp", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6941063404083252,
"train_eval_accuracy": 0.4950404594100757, "train_eval_precision":
0.49528178243774573, "train_eval_recall": 0.9952594153278904, "train_eval_f1":
0.6614159446923952, "train_eval_runtime": 36.4192,
"train_eval_samples_per_second": 210.384, "train_eval_steps_per_second": 1.647,
"train_epoch": 1.0, "val_eval_loss": 0.6973494291305542, "val_eval_accuracy":
0.46318289786223277, "val_eval_precision": 0.45933014354066987,
"val_eval_recall": 1.0, "val_eval_f1": 0.6295081967213115, "val_eval_runtime":
3.2379, "val_eval_samples_per_second": 130.022, "val_eval_steps_per_second":
1.235, "val_epoch": 1.0, "test_eval_loss": 0.6952130794525146,
"test_eval_accuracy": 0.48091603053435117, "test_eval_precision":
0.48083242059145676, "test_eval_recall": 0.9954648526077098, "test_eval_f1":
0.6484490398818316, "test_eval_runtime": 5.534, "test_eval_samples_per_second":
165.703, "test_eval_steps_per_second": 1.446, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T21:50:08.478807-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,

```

```

"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 1.287038803100586, "train_eval_accuracy":
0.4955625163142783, "train_eval_precision": 0.4955543933054393,
"train_eval_recall": 0.9981564392941796, "train_eval_f1": 0.6622979467016165,
"train_eval_runtime": 16.8721, "train_eval_samples_per_second": 454.121,
"train_eval_steps_per_second": 3.556, "train_epoch": 1.0, "val_eval_loss":
1.367303729057312, "val_eval_accuracy": 0.4584323040380047,
"val_eval_precision": 0.45714285714285713, "val_eval_recall": 1.0,
"val_eval_f1": 0.6274509803921569, "val_eval_runtime": 2.2683,
"val_eval_samples_per_second": 185.602, "val_eval_steps_per_second": 1.763,
"val_epoch": 1.0, "test_eval_loss": 1.3381584882736206, "test_eval_accuracy":
0.47546346782988, "test_eval_precision": 0.4780701754385965, "test_eval_recall":
0.9886621315192744, "test_eval_f1": 0.6444937176644494, "test_eval_runtime":
3.0284, "test_eval_samples_per_second": 302.799, "test_eval_steps_per_second":
2.642, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T21:52:17.715863-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-large", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.27.attn_norm.weight", "model.layers.27.attn.Wqkv.weight",
"model.layers.27.attn.Wo.weight", "model.layers.27.mlp_norm.weight",
"model.layers.27.mlp.Wi.weight", "model.layers.27.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 395833346, "model_trainable_params": 13309954,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,

```

```

"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-large",
"_commit_hash": "45bb4654a4d5aaff24dd11d4781fa46d39bf8c13",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 1024, "intermediate_size": 2624,
"num_hidden_layers": 28, "num_attention_heads": 16, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.7216485738754272, "train_eval_accuracy":
0.498564343513443, "train_eval_precision": 0.49675465166594546,
"train_eval_recall": 0.9070318672636292, "train_eval_f1": 0.6419384902143522,
"train_eval_runtime": 42.8424, "train_eval_samples_per_second": 178.841,
"train_eval_steps_per_second": 1.4, "train_epoch": 1.0, "val_eval_loss":
0.7473640441894531, "val_eval_accuracy": 0.44655581947743467,
"val_eval_precision": 0.4464751958224543, "val_eval_recall": 0.890625,
"val_eval_f1": 0.5947826086956521, "val_eval_runtime": 3.3663,
"val_eval_samples_per_second": 125.063, "val_eval_steps_per_second": 1.188,
"val_epoch": 1.0, "test_eval_loss": 0.7314949035644531, "test_eval_accuracy":
0.4721919302071974, "test_eval_precision": 0.474251497005988,
"test_eval_recall": 0.8979591836734694, "test_eval_f1": 0.6206896551724138,

```

```

"test_eval_runtime": 6.0258, "test_eval_samples_per_second": 152.179,
"test_eval_steps_per_second": 1.328, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T21:53:26.753781-07:00", "experiment_meta":
{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",
"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",
"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":
"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":

```

```

null, "attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":
"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":
12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,
"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.6968384981155396, "train_eval_accuracy":
0.4943878882798225, "train_eval_precision": 0.4947640418876649,
"train_eval_recall": 0.9581248353963655, "train_eval_f1": 0.6525560538116592,
"train_eval_runtime": 15.6851, "train_eval_samples_per_second": 488.49,
"train_eval_steps_per_second": 3.825, "train_epoch": 1.0, "val_eval_loss":
0.7027359008789062, "val_eval_accuracy": 0.4513064133016627,
"val_eval_precision": 0.45112781954887216, "val_eval_recall": 0.9375,
"val_eval_f1": 0.6091370558375635, "val_eval_runtime": 1.9534,
"val_eval_samples_per_second": 215.516, "val_eval_steps_per_second": 2.048,
"val_epoch": 1.0, "test_eval_loss": 0.7016682624816895, "test_eval_accuracy":
0.4820065430752454, "test_eval_precision": 0.48072562358276644,
"test_eval_recall": 0.9614512471655329, "test_eval_f1": 0.6409674981103552,
"test_eval_runtime": 2.919, "test_eval_samples_per_second": 314.154,
"test_eval_steps_per_second": 2.741, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T21:54:38.757064-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.11.rel_attn.q", "transformer.layer.11.rel_attn.k",
"transformer.layer.11.rel_attn.v", "transformer.layer.11.rel_attn.o",
"transformer.layer.11.rel_attn.r", "transformer.layer.11.rel_attn.r_r_bias",
"transformer.layer.11.rel_attn.r_s_bias",
"transformer.layer.11.rel_attn.r_w_bias",
"transformer.layer.11.rel_attn.seg_embed",
"transformer.layer.11.rel_attn.layer_norm.weight",
"transformer.layer.11.rel_attn.layer_norm.bias",
"transformer.layer.11.ff.layer_norm.weight",
"transformer.layer.11.ff.layer_norm.bias",
"transformer.layer.11.ff.layer_1.weight",
"transformer.layer.11.ff.layer_1.bias",
"transformer.layer.11.ff.layer_2.weight",
"transformer.layer.11.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 117310466, "model_trainable_params":
8270594, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 768, "n_layer": 12, "n_head": 12, "d_head": 64,
"ff_activation": "gelu", "d_inner": 3072, "untie_r": true, "attn_type": "bi",
"initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1, "mem_len":

```

```

null, "reuse_len": null, "bi_data": false, "clamp_len": -1, "same_length":
false, "summary_type": "last", "summary_use_proj": true, "summary_activation":
"tanh", "summary_last_dropout": 0.1, "start_n_top": 5, "end_n_top": 5,
"bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2, "use_mems_eval": true,
"use_mems_train": false, "return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-base-cased",
"_commit_hash": "ceaa69c7bc5e512b5007106a7ccbb7daf24b2c79",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}, {"run_metrics":
{"train_eval_loss": 0.701991856098175, "train_eval_accuracy":
0.5029663810151616, "train_eval_precision": 0.5013717421124828,
"train_eval_recall": 0.9643799472295514, "train_eval_f1": 0.6597472924187726,
"train_eval_runtime": 5.1658, "train_eval_samples_per_second": 293.659,
"train_eval_steps_per_second": 2.323, "train_epoch": 1.0, "val_eval_loss":
0.6949297189712524, "val_eval_accuracy": 0.5353535353535354,
"val_eval_precision": 0.5263157894736842, "val_eval_recall": 0.9803921568627451,
"val_eval_f1": 0.684931506849315, "val_eval_runtime": 1.8621,
"val_eval_samples_per_second": 53.165, "val_eval_steps_per_second": 0.537,
"val_epoch": 1.0, "test_eval_loss": 0.6872782707214355, "test_eval_accuracy":
0.5108695652173914, "test_eval_precision": 0.5254237288135594,
"test_eval_recall": 0.9393939393939394, "test_eval_f1": 0.6739130434782609,
"test_eval_runtime": 4.9605, "test_eval_samples_per_second": 37.093,
"test_eval_steps_per_second": 0.403, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T21:55:34.491244-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.23.rel_attn.q", "transformer.layer.23.rel_attn.k",
"transformer.layer.23.rel_attn.v", "transformer.layer.23.rel_attn.o",

```

```

"transformer.layer.23.rel_attn.r", "transformer.layer.23.rel_attn.r_r_bias",
"transformer.layer.23.rel_attn.r_s_bias",
"transformer.layer.23.rel_attn.r_w_bias",
"transformer.layer.23.rel_attn.seg_embed",
"transformer.layer.23.rel_attn.layer_norm.weight",
"transformer.layer.23.rel_attn.layer_norm.bias",
"transformer.layer.23.ff.layer_norm.weight",
"transformer.layer.23.ff.layer_norm.bias",
"transformer.layer.23.ff.layer_1.weight",
"transformer.layer.23.ff.layer_1.bias",
"transformer.layer.23.ff.layer_2.weight",
"transformer.layer.23.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 361320450, "model_trainable_params":
14697474, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 1024, "n_layer": 24, "n_head": 16, "d_head":
64, "ff_activation": "gelu", "d_inner": 4096, "untie_r": true, "attn_type":
"bi", "initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1,
"mem_len": null, "reuse_len": null, "bi_data": false, "clamp_len": -1,
"same_length": false, "summary_type": "last", "summary_use_proj": true,
"summary_activation": "tanh", "summary_last_dropout": 0.1, "start_n_top": 5,
"end_n_top": 5, "bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2,
"use_mems_eval": true, "use_mems_train": false, "return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-large-cased",
"_commit_hash": "37658b4f179eaf971d127d16bfb9ca676a93034",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}, "run_metrics":
{"train_eval_loss": 0.7027806043624878, "train_eval_accuracy":
0.5036255767963085, "train_eval_precision": 0.5067024128686327,

```

```

"train_eval_recall": 0.24934036939313983, "train_eval_f1": 0.33421750663129973,
"train_eval_runtime": 12.8571, "train_eval_samples_per_second": 117.989,
"train_eval_steps_per_second": 0.933, "train_epoch": 1.0, "val_eval_loss":
0.6968010067939758, "val_eval_accuracy": 0.5050505050505051,
"val_eval_precision": 0.5384615384615384, "val_eval_recall":
0.27450980392156865, "val_eval_f1": 0.36363636363636365, "val_eval_runtime":
2.8871, "val_eval_samples_per_second": 34.291, "val_eval_steps_per_second":
0.346, "val_epoch": 1.0, "test_eval_loss": 0.7015742659568787,
"test_eval_accuracy": 0.5108695652173914, "test_eval_precision":
0.5849056603773585, "test_eval_recall": 0.31313131313131315, "test_eval_f1":
0.40789473684210525, "test_eval_runtime": 2.5414,
"test_eval_samples_per_second": 72.4, "test_eval_steps_per_second": 0.787,
"test_epoch": 1.0}}
{"timestamp": "2025-04-10T21:57:29.687483-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.11.rel_attn.q", "transformer.layer.11.rel_attn.k",
"transformer.layer.11.rel_attn.v", "transformer.layer.11.rel_attn.o",
"transformer.layer.11.rel_attn.r", "transformer.layer.11.rel_attn.r_r_bias",
"transformer.layer.11.rel_attn.r_s_bias",
"transformer.layer.11.rel_attn.r_w_bias",
"transformer.layer.11.rel_attn.seg_embed",
"transformer.layer.11.rel_attn.layer_norm.weight",
"transformer.layer.11.rel_attn.layer_norm.bias",
"transformer.layer.11.ff.layer_norm.weight",
"transformer.layer.11.ff.layer_norm.bias",
"transformer.layer.11.ff.layer_1.weight",
"transformer.layer.11.ff.layer_1.bias",
"transformer.layer.11.ff.layer_2.weight",
"transformer.layer.11.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 117310466, "model_trainable_params":
8270594, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 768, "n_layer": 12, "n_head": 12, "d_head": 64,
"ff_activation": "gelu", "d_inner": 3072, "untie_r": true, "attn_type": "bi",
"initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1, "mem_len":
null, "reuse_len": null, "bi_data": false, "clamp_len": -1, "same_length":
false, "summary_type": "last", "summary_use_proj": true, "summary_activation":
"tanh", "summary_last_dropout": 0.1, "start_n_top": 5, "end_n_top": 5,
"bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2, "use_mems_eval": true,
"use_mems_train": false, "return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,

```



```

"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-base-cased",
"_commit_hash": "ceaa69c7bc5e512b5007106a7ccbb7daf24b2c79",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}}, "run_metrics":
{"train_eval_loss": 0.6951068639755249, "train_eval_accuracy":
0.5125293657008614, "train_eval_precision": 0.5070199275362319,
"train_eval_recall": 0.5896760600474058, "train_eval_f1": 0.5452331669304761,
"train_eval_runtime": 23.9583, "train_eval_samples_per_second": 319.805,
"train_eval_steps_per_second": 2.504, "train_epoch": 1.0, "val_eval_loss":
0.7071340680122375, "val_eval_accuracy": 0.4513064133016627,
"val_eval_precision": 0.42292490118577075, "val_eval_recall":
0.5572916666666666, "val_eval_f1": 0.48089887640449436, "val_eval_runtime":
6.534, "val_eval_samples_per_second": 64.432, "val_eval_steps_per_second":
0.612, "val_epoch": 1.0, "test_eval_loss": 0.6994464993476868,
"test_eval_accuracy": 0.5114503816793893, "test_eval_precision":
0.4934086629001883, "test_eval_recall": 0.5941043083900227, "test_eval_f1":
0.5390946502057613, "test_eval_runtime": 7.0154, "test_eval_samples_per_second":
130.712, "test_eval_steps_per_second": 1.14, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T22:00:54.769290-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.23.rel_attn.q", "transformer.layer.23.rel_attn.k",
"transformer.layer.23.rel_attn.v", "transformer.layer.23.rel_attn.o",
"transformer.layer.23.rel_attn.r", "transformer.layer.23.rel_attn.r_r_bias",
"transformer.layer.23.rel_attn.r_s_bias",
"transformer.layer.23.rel_attn.r_w_bias",
"transformer.layer.23.rel_attn.seg_embed",
"transformer.layer.23.rel_attn.layer_norm.weight",
"transformer.layer.23.rel_attn.layer_norm.bias",
"transformer.layer.23.ff.layer_norm.weight",
"transformer.layer.23.ff.layer_norm.bias",
"transformer.layer.23.ff.layer_1.weight",
"transformer.layer.23.ff.layer_1.bias",
"transformer.layer.23.ff.layer_2.weight",

```

```

"transformer.layer.23.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 361320450, "model_trainable_params":
14697474, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 1024, "n_layer": 24, "n_head": 16, "d_head":
64, "ff_activation": "gelu", "d_inner": 4096, "untie_r": true, "attn_type":
"bi", "initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1,
"mem_len": null, "reuse_len": null, "bi_data": false, "clamp_len": -1,
"same_length": false, "summary_type": "last", "summary_use_proj": true,
"summary_activation": "tanh", "summary_last_dropout": 0.1, "start_n_top": 5,
"end_n_top": 5, "bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2,
"use_mems_eval": true, "use_mems_train": false, "return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-large-cased",
"_commit_hash": "37658b4f179eaf971d127d16bfb9ca676a93034",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}}, "run_metrics":
{"train_eval_loss": 0.6971995830535889, "train_eval_accuracy":
0.49947794309579746, "train_eval_precision": 0.4971818451498072,
"train_eval_recall": 0.8828022122728469, "train_eval_f1": 0.6361134832526805,
"train_eval_runtime": 70.2669, "train_eval_samples_per_second": 109.041,
"train_eval_steps_per_second": 0.854, "train_epoch": 1.0, "val_eval_loss":
0.7058453559875488, "val_eval_accuracy": 0.42992874109263657,
"val_eval_precision": 0.4344262295081967, "val_eval_recall": 0.828125,
"val_eval_f1": 0.5698924731182796, "val_eval_runtime": 15.0527,
"val_eval_samples_per_second": 27.968, "val_eval_steps_per_second": 0.266,
"val_epoch": 1.0, "test_eval_loss": 0.6994321942329407, "test_eval_accuracy":
0.48527808069792805, "test_eval_precision": 0.48098159509202454,
"test_eval_recall": 0.8888888888888888, "test_eval_f1": 0.6242038216560509,
"test_eval_runtime": 15.3381, "test_eval_samples_per_second": 59.786,

```

```

"test_eval_steps_per_second": 0.522, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T22:02:00.151139-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaeff65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.7204152345657349,
"train_eval_accuracy": 0.5280158206987475, "train_eval_precision":
0.5580110497237569, "train_eval_recall": 0.26649076517150394, "train_eval_f1":
0.3607142857142857, "train_eval_runtime": 8.0012,
"train_eval_samples_per_second": 189.596, "train_eval_steps_per_second": 1.5,
"train_epoch": 1.0, "val_eval_loss": 0.762324333190918, "val_eval_accuracy":
0.4444444444444444, "val_eval_precision": 0.4333333333333333,
"val_eval_recall": 0.2549019607843137, "val_eval_f1": 0.32098765432098764,
"val_eval_runtime": 1.4804, "val_eval_samples_per_second": 66.876,

```

```

"val_eval_steps_per_second": 0.676, "val_epoch": 1.0, "test_eval_loss":
0.7283492684364319, "test_eval_accuracy": 0.5163043478260869,
"test_eval_precision": 0.6086956521739131, "test_eval_recall":
0.2828282828282828, "test_eval_f1": 0.38620689655172413, "test_eval_runtime":
2.9288, "test_eval_samples_per_second": 62.825, "test_eval_steps_per_second":
0.683, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T22:02:42.050945-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,

```

```

"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdp", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6987705230712891,
"train_eval_accuracy": 0.4996704021094265, "train_eval_precision":
0.4996704021094265, "train_eval_recall": 1.0, "train_eval_f1":
0.6663736263736264, "train_eval_runtime": 8.5098,
"train_eval_samples_per_second": 178.264, "train_eval_steps_per_second": 1.41,
"train_epoch": 1.0, "val_eval_loss": 0.6962690949440002, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 2.1027,
"val_eval_samples_per_second": 47.081, "val_eval_steps_per_second": 0.476,
"val_epoch": 1.0, "test_eval_loss": 0.6907474398612976, "test_eval_accuracy":
0.5380434782608695, "test_eval_precision": 0.5380434782608695,
"test_eval_recall": 1.0, "test_eval_f1": 0.6996466431095406,
"test_eval_runtime": 2.0999, "test_eval_samples_per_second": 87.625,
"test_eval_steps_per_second": 0.952, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T22:03:28.917885-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,

```

```

"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 2.1026854515075684, "train_eval_accuracy":
0.4996704021094265, "train_eval_precision": 0.4996704021094265,
"train_eval_recall": 1.0, "train_eval_f1": 0.6663736263736264,
"train_eval_runtime": 4.3024, "train_eval_samples_per_second": 352.594,
"train_eval_steps_per_second": 2.789, "train_epoch": 1.0, "val_eval_loss":
2.0287740230560303, "val_eval_accuracy": 0.5151515151515151,
"val_eval_precision": 0.5151515151515151, "val_eval_recall": 1.0, "val_eval_f1":
0.68, "val_eval_runtime": 2.6817, "val_eval_samples_per_second": 36.917,
"val_eval_steps_per_second": 0.373, "val_epoch": 1.0, "test_eval_loss":
1.95679771900177, "test_eval_accuracy": 0.5380434782608695,
"test_eval_precision": 0.5380434782608695, "test_eval_recall": 1.0,
"test_eval_f1": 0.6996466431095406, "test_eval_runtime": 2.6347,
"test_eval_samples_per_second": 69.838, "test_eval_steps_per_second": 0.759,
"test_epoch": 1.0}}
{"timestamp": "2025-04-10T22:04:23.715015-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-large", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.27.attn_norm.weight", "model.layers.27.attn.Wqkv.weight",
"model.layers.27.attn.Wo.weight", "model.layers.27.mlp_norm.weight",
"model.layers.27.mlp.Wi.weight", "model.layers.27.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 395833346, "model_trainable_params": 13309954,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,

```

```

"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-large",
"_commit_hash": "45bb4654a4d5aaff24dd11d4781fa46d39bf8c13",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 1024, "intermediate_size": 2624,
"num_hidden_layers": 28, "num_attention_heads": 16, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 1.3652886152267456, "train_eval_accuracy":
0.5003295978905735, "train_eval_precision": 0.0, "train_eval_recall": 0.0,
"train_eval_f1": 0.0, "train_eval_runtime": 11.1025,
"train_eval_samples_per_second": 136.636, "train_eval_steps_per_second": 1.081,
"train_epoch": 1.0, "val_eval_loss": 1.3708757162094116, "val_eval_accuracy":
0.48484848484848486, "val_eval_precision": 0.0, "val_eval_recall": 0.0,
"val_eval_f1": 0.0, "val_eval_runtime": 2.9017, "val_eval_samples_per_second":
34.118, "val_eval_steps_per_second": 0.345, "val_epoch": 1.0, "test_eval_loss":
1.411220669746399, "test_eval_accuracy": 0.46195652173913043,
"test_eval_precision": 0.0, "test_eval_recall": 0.0, "test_eval_f1": 0.0,
"test_eval_runtime": 2.6697, "test_eval_samples_per_second": 68.922,
"test_eval_steps_per_second": 0.749, "test_epoch": 1.0}}

```

```

{"timestamp": "2025-04-10T22:04:51.408883-07:00", "experiment_meta":
{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",
"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",
"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":
"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":
null, "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":

```



```

"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":
12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,
"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.6976747512817383, "train_eval_accuracy":
0.46868820039551745, "train_eval_precision": 0.45, "train_eval_recall":
0.2849604221635884, "train_eval_f1": 0.34894991922455576, "train_eval_runtime":
4.1494, "train_eval_samples_per_second": 365.594, "train_eval_steps_per_second":
2.892, "train_epoch": 1.0, "val_eval_loss": 0.6982458233833313,
"val_eval_accuracy": 0.43434343434343436, "val_eval_precision":
0.42424242424242425, "val_eval_recall": 0.27450980392156865, "val_eval_f1":
0.3333333333333333, "val_eval_runtime": 1.4629, "val_eval_samples_per_second":
67.674, "val_eval_steps_per_second": 0.684, "val_epoch": 1.0, "test_eval_loss":
0.6952725052833557, "test_eval_accuracy": 0.5271739130434783,
"test_eval_precision": 0.59375, "test_eval_recall": 0.3838383838383838,
"test_eval_f1": 0.4662576687116564, "test_eval_runtime": 1.5748,
"test_eval_samples_per_second": 116.842, "test_eval_steps_per_second": 1.27,
"test_epoch": 1.0}}
{"timestamp": "2025-04-10T22:11:39.458805-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.11.rel_attn.q", "transformer.layer.11.rel_attn.k",
"transformer.layer.11.rel_attn.v", "transformer.layer.11.rel_attn.o",
"transformer.layer.11.rel_attn.r", "transformer.layer.11.rel_attn.r_r_bias",
"transformer.layer.11.rel_attn.r_s_bias",
"transformer.layer.11.rel_attn.r_w_bias",
"transformer.layer.11.rel_attn.seg_embed",
"transformer.layer.11.rel_attn.layer_norm.weight",
"transformer.layer.11.rel_attn.layer_norm.bias",
"transformer.layer.11.ff.layer_norm.weight",
"transformer.layer.11.ff.layer_norm.bias",
"transformer.layer.11.ff.layer_1.weight",
"transformer.layer.11.ff.layer_1.bias",
"transformer.layer.11.ff.layer_2.weight",
"transformer.layer.11.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 117310466, "model_trainable_params":
8270594, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 768, "n_layer": 12, "n_head": 12, "d_head": 64,
"ff_activation": "gelu", "d_inner": 3072, "untie_r": true, "attn_type": "bi",
"initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1, "mem_len":
null, "reuse_len": null, "bi_data": false, "clamp_len": -1, "same_length":
false, "summary_type": "last", "summary_use_proj": true, "summary_activation":

```

```

"tanh", "summary_last_dropout": 0.1, "start_n_top": 5, "end_n_top": 5,
"bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2, "use_mems_eval": true,
"use_mems_train": false, "return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-base-cased",
"_commit_hash": "ceaa69c7bc5e512b5007106a7ccbb7daf24b2c79",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}, {"run_metrics":
{"train_eval_loss": 0.7059321403503418, "train_eval_accuracy":
0.4963744232036915, "train_eval_precision": 0.4978386167146974,
"train_eval_recall": 0.9116094986807388, "train_eval_f1": 0.6439888164026095,
"train_eval_runtime": 5.051, "train_eval_samples_per_second": 300.334,
"train_eval_steps_per_second": 2.376, "train_epoch": 1.0, "val_eval_loss":
0.7009491324424744, "val_eval_accuracy": 0.5151515151515151,
"val_eval_precision": 0.5164835164835165, "val_eval_recall": 0.9215686274509803,
"val_eval_f1": 0.6619718309859155, "val_eval_runtime": 2.3901,
"val_eval_samples_per_second": 41.42, "val_eval_steps_per_second": 0.418,
"val_epoch": 1.0, "test_eval_loss": 0.6971263289451599, "test_eval_accuracy":
0.5108695652173914, "test_eval_precision": 0.5266272189349113,
"test_eval_recall": 0.898989898989899, "test_eval_f1": 0.664179104477612,
"test_eval_runtime": 2.0259, "test_eval_samples_per_second": 90.824,
"test_eval_steps_per_second": 0.987, "test_epoch": 1.0}}
{"timestamp": "2025-04-10T22:12:32.298158-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.23.rel_attn.q", "transformer.layer.23.rel_attn.k",
"transformer.layer.23.rel_attn.v", "transformer.layer.23.rel_attn.o",
"transformer.layer.23.rel_attn.r", "transformer.layer.23.rel_attn.r_r_bias",
"transformer.layer.23.rel_attn.r_s_bias",

```

```

"transformer.layer.23.rel_attn.r_w_bias",
"transformer.layer.23.rel_attn.seg_embed",
"transformer.layer.23.rel_attn.layer_norm.weight",
"transformer.layer.23.rel_attn.layer_norm.bias",
"transformer.layer.23.ff.layer_norm.weight",
"transformer.layer.23.ff.layer_norm.bias",
"transformer.layer.23.ff.layer_1.weight",
"transformer.layer.23.ff.layer_1.bias",
"transformer.layer.23.ff.layer_2.weight",
"transformer.layer.23.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 361320450, "model_trainable_params":
14697474, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 1024, "n_layer": 24, "n_head": 16, "d_head":
64, "ff_activation": "gelu", "d_inner": 4096, "untie_r": true, "attn_type":
"bi", "initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1,
"mem_len": null, "reuse_len": null, "bi_data": false, "clamp_len": -1,
"same_length": false, "summary_type": "last", "summary_use_proj": true,
"summary_activation": "tanh", "summary_last_dropout": 0.1, "start_n_top": 5,
"end_n_top": 5, "bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2,
"use_mems_eval": true, "use_mems_train": false, "return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-large-cased",
"_commit_hash": "37658b4f179eaf971d127d16bfb9ca676a93034",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}, "run_metrics":
{"train_eval_loss": 0.6988334655761719, "train_eval_accuracy":
0.4897824653922215, "train_eval_precision": 0.4913978494623656,
"train_eval_recall": 0.6029023746701847, "train_eval_f1": 0.5414691943127962,
"train_eval_runtime": 12.3141, "train_eval_samples_per_second": 123.192,

```

```

"train_eval_steps_per_second": 0.974, "train_epoch": 1.0, "val_eval_loss":
0.7020902633666992, "val_eval_accuracy": 0.48484848484848486,
"val_eval_precision": 0.5, "val_eval_recall": 0.6078431372549019, "val_eval_f1":
0.5486725663716814, "val_eval_runtime": 2.5747, "val_eval_samples_per_second":
38.452, "val_eval_steps_per_second": 0.388, "val_epoch": 1.0, "test_eval_loss":
0.697321355342865, "test_eval_accuracy": 0.483695652173913,
"test_eval_precision": 0.5185185185185185, "test_eval_recall":
0.5656565656565656, "test_eval_f1": 0.5410628019323671, "test_eval_runtime":
3.222, "test_eval_samples_per_second": 57.108, "test_eval_steps_per_second":
0.621, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T05:08:44.033681-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6971537470817566, "train_eval_accuracy": 0.49947794309579746,
"train_eval_precision": 0.49585876198779427, "train_eval_recall":

```

```

0.599157229391625, "train_eval_f1": 0.5426356589147286, "train_eval_runtime":
16.5245, "train_eval_samples_per_second": 463.675,
"train_eval_steps_per_second": 3.631, "train_epoch": 1.0, "val_eval_loss":
0.696948766708374, "val_eval_accuracy": 0.5201900237529691,
"val_eval_precision": 0.4791666666666667, "val_eval_recall": 0.5989583333333334,
"val_eval_f1": 0.5324074074074074, "val_eval_runtime": 5.4029,
"val_eval_samples_per_second": 77.921, "val_eval_steps_per_second": 0.74,
"val_epoch": 1.0, "test_eval_loss": 0.6963955760002136, "test_eval_accuracy":
0.49727371864776443, "test_eval_precision": 0.48161764705882354,
"test_eval_recall": 0.5941043083900227, "test_eval_f1": 0.531979695431472,
"test_eval_runtime": 6.8485, "test_eval_samples_per_second": 133.899,
"test_eval_steps_per_second": 1.168, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T05:09:53.794719-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,

```

```

"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6931397914886475,
"train_eval_accuracy": 0.49569303054032887, "train_eval_precision":
0.4954502240934402, "train_eval_recall": 0.9607584935475375, "train_eval_f1":
0.6537634408602151, "train_eval_runtime": 16.0572,
"train_eval_samples_per_second": 477.17, "train_eval_steps_per_second": 3.737,
"train_epoch": 1.0, "val_eval_loss": 0.6936591863632202, "val_eval_accuracy":
0.4513064133016627, "val_eval_precision": 0.45161290322580644,
"val_eval_recall": 0.9479166666666666, "val_eval_f1": 0.611764705882353,
"val_eval_runtime": 5.3734, "val_eval_samples_per_second": 78.349,
"val_eval_steps_per_second": 0.744, "val_epoch": 1.0, "test_eval_loss":
0.6932082176208496, "test_eval_accuracy": 0.48636859323882226,
"test_eval_precision": 0.48299319727891155, "test_eval_recall":
0.9659863945578231, "test_eval_f1": 0.6439909297052154, "test_eval_runtime":
6.1841, "test_eval_samples_per_second": 148.284, "test_eval_steps_per_second":
1.294, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T05:13:18.402713-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,

```

```

"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaeff65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6956863403320312,
"train_eval_accuracy": 0.5007830853563039, "train_eval_precision":
0.4964249233912155, "train_eval_recall": 0.5119831445878325, "train_eval_f1":
0.5040840140023337, "train_eval_runtime": 39.5427,
"train_eval_samples_per_second": 193.765, "train_eval_steps_per_second": 1.517,
"train_epoch": 1.0, "val_eval_loss": 0.698200523853302, "val_eval_accuracy":
0.4750593824228028, "val_eval_precision": 0.43612334801762115,
"val_eval_recall": 0.515625, "val_eval_f1": 0.47255369928400953,
"val_eval_runtime": 6.7175, "val_eval_samples_per_second": 62.672,
"val_eval_steps_per_second": 0.595, "val_epoch": 1.0, "test_eval_loss":
0.6970794200897217, "test_eval_accuracy": 0.4798255179934569,
"test_eval_precision": 0.4608695652173913, "test_eval_recall":
0.48072562358276644, "test_eval_f1": 0.47058823529411764, "test_eval_runtime":
8.9145, "test_eval_samples_per_second": 102.866, "test_eval_steps_per_second":
0.897, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T05:15:48.159564-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",

```

```

"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6936471462249756,
"train_eval_accuracy": 0.4969981728008353, "train_eval_precision":
0.49604770489529887, "train_eval_recall": 0.9420595206742165, "train_eval_f1":
0.649890988372093, "train_eval_runtime": 41.914,
"train_eval_samples_per_second": 182.803, "train_eval_steps_per_second": 1.432,
"train_epoch": 1.0, "val_eval_loss": 0.6961225271224976, "val_eval_accuracy":
0.45605700712589076, "val_eval_precision": 0.45363408521303256,
"val_eval_recall": 0.9427083333333334, "val_eval_f1": 0.6125211505922166,

```



```

"val_eval_runtime": 8.9308, "val_eval_samples_per_second": 47.14,
"val_eval_steps_per_second": 0.448, "val_epoch": 1.0, "test_eval_loss":
0.6944529414176941, "test_eval_accuracy": 0.48964013086150493,
"test_eval_precision": 0.48413631022326675, "test_eval_recall":
0.9342403628117913, "test_eval_f1": 0.6377708978328174, "test_eval_runtime":
12.5977, "test_eval_samples_per_second": 72.791, "test_eval_steps_per_second":
0.635, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T05:17:36.132486-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,

```

```

"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.863323450088501, "train_eval_accuracy":
0.4973897154789872, "train_eval_precision": 0.4962290502793296,
"train_eval_recall": 0.9357387411114038, "train_eval_f1": 0.6485351829880441,
"train_eval_runtime": 21.6026, "train_eval_samples_per_second": 354.68,
"train_eval_steps_per_second": 2.777, "train_epoch": 1.0, "val_eval_loss":
0.890388548374176, "val_eval_accuracy": 0.4489311163895487,
"val_eval_precision": 0.4502487562189055, "val_eval_recall": 0.9427083333333334,
"val_eval_f1": 0.6094276094276094, "val_eval_runtime": 5.8647,
"val_eval_samples_per_second": 71.786, "val_eval_steps_per_second": 0.682,
"val_epoch": 1.0, "test_eval_loss": 0.8967874050140381, "test_eval_accuracy":
0.4798255179934569, "test_eval_precision": 0.47897196261682246,
"test_eval_recall": 0.9297052154195011, "test_eval_f1": 0.6322282189668466,
"test_eval_runtime": 9.505, "test_eval_samples_per_second": 96.476,
"test_eval_steps_per_second": 0.842, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T05:20:10.800671-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-large", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.27.attn_norm.weight", "model.layers.27.attn.Wqkv.weight",
"model.layers.27.attn.Wo.weight", "model.layers.27.mlp_norm.weight",
"model.layers.27.mlp.Wi.weight", "model.layers.27.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 395833346, "model_trainable_params": 13309954,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":

```

```

50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-large",
"_commit_hash": "45bb4654a4d5aaff24dd11d4781fa46d39bf8c13",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 1024, "intermediate_size": 2624,
"num_hidden_layers": 28, "num_attention_heads": 16, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.6994955539703369, "train_eval_accuracy":
0.506525711302532, "train_eval_precision": 0.501719690455718,
"train_eval_recall": 0.6146958124835397, "train_eval_f1": 0.5524914191028524,
"train_eval_runtime": 46.6534, "train_eval_samples_per_second": 164.232,
"train_eval_steps_per_second": 1.286, "train_epoch": 1.0, "val_eval_loss":
0.7122632265090942, "val_eval_accuracy": 0.4845605700712589,
"val_eval_precision": 0.4517374517374517, "val_eval_recall": 0.609375,
"val_eval_f1": 0.5188470066518847, "val_eval_runtime": 7.0676,
"val_eval_samples_per_second": 59.568, "val_eval_steps_per_second": 0.566,
"val_epoch": 1.0, "test_eval_loss": 0.7037101984024048, "test_eval_accuracy":
0.5005452562704471, "test_eval_precision": 0.4847942754919499,
"test_eval_recall": 0.6145124716553289, "test_eval_f1": 0.542,
"test_eval_runtime": 9.7239, "test_eval_samples_per_second": 94.304,
"test_eval_steps_per_second": 0.823, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T05:21:45.162764-07:00", "experiment_meta":
{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",

```

```

"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",
"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":
"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":
null, "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":
"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":
12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,
"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.6978607773780823, "train_eval_accuracy":
0.493474288697468, "train_eval_precision": 0.4944223107569721,
"train_eval_recall": 0.9805109296813274, "train_eval_f1": 0.6573673523439569,
"train_eval_runtime": 19.3132, "train_eval_samples_per_second": 396.724,
"train_eval_steps_per_second": 3.107, "train_epoch": 1.0, "val_eval_loss":
0.7042899131774902, "val_eval_accuracy": 0.45605700712589076,
"val_eval_precision": 0.45498783454987834, "val_eval_recall":

```

```

0.9739583333333334, "val_eval_f1": 0.6202321724709784, "val_eval_runtime":
5.7063, "val_eval_samples_per_second": 73.778, "val_eval_steps_per_second":
0.701, "val_epoch": 1.0, "test_eval_loss": 0.7036487460136414,
"test_eval_accuracy": 0.4798255179934569, "test_eval_precision":
0.4799554565701559, "test_eval_recall": 0.9773242630385488, "test_eval_f1":
0.643764002987304, "test_eval_runtime": 6.4848, "test_eval_samples_per_second":
141.408, "test_eval_steps_per_second": 1.234, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T05:24:38.070986-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.11.rel_attn.q", "transformer.layer.11.rel_attn.k",
"transformer.layer.11.rel_attn.v", "transformer.layer.11.rel_attn.o",
"transformer.layer.11.rel_attn.r", "transformer.layer.11.rel_attn.r_r_bias",
"transformer.layer.11.rel_attn.r_s_bias",
"transformer.layer.11.rel_attn.r_w_bias",
"transformer.layer.11.rel_attn.seg_embed",
"transformer.layer.11.rel_attn.layer_norm.weight",
"transformer.layer.11.rel_attn.layer_norm.bias",
"transformer.layer.11.ff.layer_norm.weight",
"transformer.layer.11.ff.layer_norm.bias",
"transformer.layer.11.ff.layer_1.weight",
"transformer.layer.11.ff.layer_1.bias",
"transformer.layer.11.ff.layer_2.weight",
"transformer.layer.11.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 117310466, "model_trainable_params":
8270594, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 768, "n_layer": 12, "n_head": 12, "d_head": 64,
"ff_activation": "gelu", "d_inner": 3072, "untie_r": true, "attn_type": "bi",
"initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1, "mem_len":
null, "reuse_len": null, "bi_data": false, "clamp_len": -1, "same_length":
false, "summary_type": "last", "summary_use_proj": true, "summary_activation":
"tanh", "summary_last_dropout": 0.1, "start_n_top": 5, "end_n_top": 5,
"bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2, "use_mems_eval": true,
"use_mems_train": false, "return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,

```

```

"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-base-cased",
"_commit_hash": "ceaa69c7bc5e512b5007106a7ccbb7daf24b2c79",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}}, "run_metrics":
{"train_eval_loss": 0.6965945959091187, "train_eval_accuracy":
0.49477943095797444, "train_eval_precision": 0.4930003783579266,
"train_eval_recall": 0.6863313141954175, "train_eval_f1": 0.5738192227237697,
"train_eval_runtime": 24.2726, "train_eval_samples_per_second": 315.665,
"train_eval_steps_per_second": 2.472, "train_epoch": 1.0, "val_eval_loss":
0.7004374265670776, "val_eval_accuracy": 0.4679334916864608,
"val_eval_precision": 0.4444444444444444, "val_eval_recall": 0.6666666666666666,
"val_eval_f1": 0.5333333333333333, "val_eval_runtime": 5.696,
"val_eval_samples_per_second": 73.911, "val_eval_steps_per_second": 0.702,
"val_epoch": 1.0, "test_eval_loss": 0.6975250840187073, "test_eval_accuracy":
0.48964013086150493, "test_eval_precision": 0.47938931297709925,
"test_eval_recall": 0.7120181405895691, "test_eval_f1": 0.572992700729927,
"test_eval_runtime": 6.9282, "test_eval_samples_per_second": 132.357,
"test_eval_steps_per_second": 1.155, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T05:27:38.041818-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"pos_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.23.rel_attn.q", "transformer.layer.23.rel_attn.k",
"transformer.layer.23.rel_attn.v", "transformer.layer.23.rel_attn.o",
"transformer.layer.23.rel_attn.r", "transformer.layer.23.rel_attn.r_r_bias",
"transformer.layer.23.rel_attn.r_s_bias",
"transformer.layer.23.rel_attn.r_w_bias",
"transformer.layer.23.rel_attn.seg_embed",
"transformer.layer.23.rel_attn.layer_norm.weight",
"transformer.layer.23.rel_attn.layer_norm.bias",
"transformer.layer.23.ff.layer_norm.weight",
"transformer.layer.23.ff.layer_norm.bias",
"transformer.layer.23.ff.layer_1.weight",
"transformer.layer.23.ff.layer_1.bias",
"transformer.layer.23.ff.layer_2.weight",
"transformer.layer.23.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]}},
"model_details": {"model_total_params": 361320450, "model_trainable_params":
14697474, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 1024, "n_layer": 24, "n_head": 16, "d_head":
64, "ff_activation": "gelu", "d_inner": 4096, "untie_r": true, "attn_type":

```

```

"bi", "initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1,
"mem_len": null, "reuse_len": null, "bi_data": false, "clamp_len": -1,
"same_length": false, "summary_type": "last", "summary_use_proj": true,
"summary_activation": "tanh", "summary_last_dropout": 0.1, "start_n_top": 5,
"end_n_top": 5, "bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2,
"use_mems_eval": true, "use_mems_train": false, "return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-large-cased",
"_commit_hash": "37658b4f179eaf971d127d16bfb9ca676a93034",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}, "run_metrics":
{"train_eval_loss": 0.6975764036178589, "train_eval_accuracy":
0.5019577133907596, "train_eval_precision": 0.49860232455495074,
"train_eval_recall": 0.8925467474321833, "train_eval_f1": 0.6397961110062299,
"train_eval_runtime": 57.175, "train_eval_samples_per_second": 134.01,
"train_eval_steps_per_second": 1.049, "train_epoch": 1.0, "val_eval_loss":
0.7078699469566345, "val_eval_accuracy": 0.43705463182897863,
"val_eval_precision": 0.43967828418230565, "val_eval_recall":
0.8541666666666666, "val_eval_f1": 0.5805309734513274, "val_eval_runtime":
7.5915, "val_eval_samples_per_second": 55.457, "val_eval_steps_per_second":
0.527, "val_epoch": 1.0, "test_eval_loss": 0.6995231509208679,
"test_eval_accuracy": 0.48964013086150493, "test_eval_precision":
0.48375451263537905, "test_eval_recall": 0.9115646258503401, "test_eval_f1":
0.6320754716981132, "test_eval_runtime": 10.9753,
"test_eval_samples_per_second": 83.551, "test_eval_steps_per_second": 0.729,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T05:33:28.481648-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":

```

```

["bert.encoder.layer.11.", "bert.pooler.", "classifier."]], "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
1.0716125965118408, "train_eval_accuracy": 0.5003295978905735,
"train_eval_precision": 0.0, "train_eval_recall": 0.0, "train_eval_f1": 0.0,
"train_eval_runtime": 3.504, "train_eval_samples_per_second": 432.932,
"train_eval_steps_per_second": 3.425, "train_epoch": 1.0, "val_eval_loss":
1.0928629636764526, "val_eval_accuracy": 0.48484848484848486,
"val_eval_precision": 0.0, "val_eval_recall": 0.0, "val_eval_f1": 0.0,
"val_eval_runtime": 1.1376, "val_eval_samples_per_second": 87.026,
"val_eval_steps_per_second": 0.879, "val_epoch": 1.0, "test_eval_loss":
1.1112936735153198, "test_eval_accuracy": 0.46195652173913043,
"test_eval_precision": 0.0, "test_eval_recall": 0.0, "test_eval_f1": 0.0,
"test_eval_runtime": 1.3462, "test_eval_samples_per_second": 136.681,
"test_eval_steps_per_second": 1.486, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T05:33:48.519010-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":

```



```
[
  "roberta.encoder.layer.11.attention.self.query.weight",
  "roberta.encoder.layer.11.attention.self.query.bias",
  "roberta.encoder.layer.11.attention.self.key.weight",
  "roberta.encoder.layer.11.attention.self.key.bias",
  "roberta.encoder.layer.11.attention.self.value.weight",
  "roberta.encoder.layer.11.attention.self.value.bias",
  "roberta.encoder.layer.11.attention.output.dense.weight",
  "roberta.encoder.layer.11.attention.output.dense.bias",
  "roberta.encoder.layer.11.attention.output.LayerNorm.weight",
  "roberta.encoder.layer.11.attention.output.LayerNorm.bias",
  "roberta.encoder.layer.11.intermediate.dense.weight",
  "roberta.encoder.layer.11.intermediate.dense.bias",
  "roberta.encoder.layer.11.output.dense.weight",
  "roberta.encoder.layer.11.output.dense.bias",
  "roberta.encoder.layer.11.output.LayerNorm.weight",
  "roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
  "classifier.dense.bias", "classifier.out_proj.weight",
  "classifier.out_proj.bias"]}, {"model_details": {"model_total_params": 124647170,
  "model_trainable_params": 7680002, "pooler_activation_function": "N/A",
  "config_attributes": {"return_dict": true, "output_hidden_states": false,
  "output_attentions": false, "torchscript": false, "torch_dtype": "float32",
  "use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
  "tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
  false, "is_decoder": false, "cross_attention_hidden_size": null,
  "add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
  "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
  "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
  "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
  1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
  "bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
  "return_dict_in_generate": false, "forced_bos_token_id": null,
  "forced_eos_token_id": null, "remove_invalid_values": false,
  "exponential_decay_length_penalty": null, "suppress_tokens": null,
  "begin_suppress_tokens": null, "architectures":
  ["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
  "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
  "tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
  "eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
  "task_specific_params": null, "problem_type": "single_label_classification",
  "_name_or_path": "roberta-base", "_commit_hash":
  "e2da8e2f811d1448a5b465c236feacd80ffb7b", "_attn_implementation_internal":
  "sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
  "model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
  "num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
  "intermediate_size": 3072, "hidden_dropout_prob": 0.1,
  "attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
  "type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
  "position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
```

```

null}}, "run_metrics": {"train_eval_loss": 0.6938470005989075,
"train_eval_accuracy": 0.4996704021094265, "train_eval_precision":
0.4996704021094265, "train_eval_recall": 1.0, "train_eval_f1":
0.6663736263736264, "train_eval_runtime": 3.4943,
"train_eval_samples_per_second": 434.138, "train_eval_steps_per_second": 3.434,
"train_epoch": 1.0, "val_eval_loss": 0.6930067539215088, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 1.4275,
"val_eval_samples_per_second": 69.351, "val_eval_steps_per_second": 0.701,
"val_epoch": 1.0, "test_eval_loss": 0.6914202570915222, "test_eval_accuracy":
0.5380434782608695, "test_eval_precision": 0.5380434782608695,
"test_eval_recall": 1.0, "test_eval_f1": 0.6996466431095406,
"test_eval_runtime": 1.4852, "test_eval_samples_per_second": 123.891,
"test_eval_steps_per_second": 1.347, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T05:39:46.289282-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaeff65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",

```

```

"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6944723129272461,
"train_eval_accuracy": 0.5201054713249835, "train_eval_precision":
0.556390977443609, "train_eval_recall": 0.19525065963060687, "train_eval_f1":
0.2890625, "train_eval_runtime": 8.6231, "train_eval_samples_per_second":
175.924, "train_eval_steps_per_second": 1.392, "train_epoch": 1.0,
"val_eval_loss": 0.7147032022476196, "val_eval_accuracy": 0.46464646464646464,
"val_eval_precision": 0.44444444444444444, "val_eval_recall": 0.1568627450980392,
"val_eval_f1": 0.2318840579710145, "val_eval_runtime": 2.7155,
"val_eval_samples_per_second": 36.458, "val_eval_steps_per_second": 0.368,
"val_epoch": 1.0, "test_eval_loss": 0.7009161114692688, "test_eval_accuracy":
0.5, "test_eval_precision": 0.6206896551724138, "test_eval_recall":
0.18181818181818182, "test_eval_f1": 0.28125, "test_eval_runtime": 2.939,
"test_eval_samples_per_second": 62.606, "test_eval_steps_per_second": 0.681,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T05:40:24.231462-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,

```

```

"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6945023536682129,
"train_eval_accuracy": 0.4996704021094265, "train_eval_precision":
0.4996704021094265, "train_eval_recall": 1.0, "train_eval_f1":
0.6663736263736264, "train_eval_runtime": 7.9998,
"train_eval_samples_per_second": 189.631, "train_eval_steps_per_second": 1.5,
"train_epoch": 1.0, "val_eval_loss": 0.6943785548210144, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 2.1023,
"val_eval_samples_per_second": 47.092, "val_eval_steps_per_second": 0.476,
"val_epoch": 1.0, "test_eval_loss": 0.691678524017334, "test_eval_accuracy":
0.5380434782608695, "test_eval_precision": 0.5380434782608695,
"test_eval_recall": 1.0, "test_eval_f1": 0.6996466431095406,
"test_eval_runtime": 2.2696, "test_eval_samples_per_second": 81.073,
"test_eval_steps_per_second": 0.881, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T05:41:06.994680-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,

```

```

"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rop_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rop_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 2.2697854042053223, "train_eval_accuracy":
0.4976928147659855, "train_eval_precision": 0.4986754966887417,
"train_eval_recall": 0.9934036939313984, "train_eval_f1": 0.6640211640211641,
"train_eval_runtime": 4.2872, "train_eval_samples_per_second": 353.843,
"train_eval_steps_per_second": 2.799, "train_epoch": 1.0, "val_eval_loss":
2.3852405548095703, "val_eval_accuracy": 0.5151515151515151,
"val_eval_precision": 0.5151515151515151, "val_eval_recall": 1.0, "val_eval_f1":
0.68, "val_eval_runtime": 1.1834, "val_eval_samples_per_second": 83.658,
"val_eval_steps_per_second": 0.845, "val_epoch": 1.0, "test_eval_loss":
2.0180070400238037, "test_eval_accuracy": 0.5489130434782609,
"test_eval_precision": 0.5439560439560439, "test_eval_recall": 1.0,
"test_eval_f1": 0.7046263345195729, "test_eval_runtime": 1.8191,
"test_eval_samples_per_second": 101.148, "test_eval_steps_per_second": 1.099,

```

```

"test_epoch": 1.0}}
{"timestamp": "2025-04-11T05:42:03.435442-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-large", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.27.attn_norm.weight", "model.layers.27.attn.Wqkv.weight",
"model.layers.27.attn.Wo.weight", "model.layers.27.mlp_norm.weight",
"model.layers.27.mlp.Wi.weight", "model.layers.27.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 395833346, "model_trainable_params": 13309954,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-large",
"_commit_hash": "45bb4654a4d5aaff24dd11d4781fa46d39bf8c13",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 1024, "intermediate_size": 2624,
"num_hidden_layers": 28, "num_attention_heads": 16, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":

```

```

{"train_eval_loss": 0.7780663371086121, "train_eval_accuracy":
0.5234014502307185, "train_eval_precision": 0.5124732715609408,
"train_eval_recall": 0.9485488126649076, "train_eval_f1": 0.6654326700601574,
"train_eval_runtime": 9.6398, "train_eval_samples_per_second": 157.368,
"train_eval_steps_per_second": 1.245, "train_epoch": 1.0, "val_eval_loss":
0.7811765670776367, "val_eval_accuracy": 0.5656565656565656,
"val_eval_precision": 0.5425531914893617, "val_eval_recall": 1.0, "val_eval_f1":
0.7034482758620689, "val_eval_runtime": 1.854, "val_eval_samples_per_second":
53.398, "val_eval_steps_per_second": 0.539, "val_epoch": 1.0, "test_eval_loss":
0.8249892592430115, "test_eval_accuracy": 0.5217391304347826,
"test_eval_precision": 0.5314285714285715, "test_eval_recall":
0.9393939393939394, "test_eval_f1": 0.6788321167883211, "test_eval_runtime":
2.6491, "test_eval_samples_per_second": 69.458, "test_eval_steps_per_second":
0.755, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T05:42:29.296539-07:00", "experiment_meta":
{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",
"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",
"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,

```

```

"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":
"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":
null, "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":
"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":
12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,
"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.6939409971237183, "train_eval_accuracy":
0.5088991430454846, "train_eval_precision": 0.524904214559387,
"train_eval_recall": 0.18073878627968337, "train_eval_f1": 0.2688910696761531,
"train_eval_runtime": 4.1445, "train_eval_samples_per_second": 366.023,
"train_eval_steps_per_second": 2.895, "train_epoch": 1.0, "val_eval_loss":
0.6993660926818848, "val_eval_accuracy": 0.5050505050505051,
"val_eval_precision": 0.5833333333333334, "val_eval_recall":
0.13725490196078433, "val_eval_f1": 0.2222222222222222, "val_eval_runtime":
1.5129, "val_eval_samples_per_second": 65.435, "val_eval_steps_per_second":
0.661, "val_epoch": 1.0, "test_eval_loss": 0.7026337385177612,
"test_eval_accuracy": 0.483695652173913, "test_eval_precision":
0.5833333333333334, "test_eval_recall": 0.1414141414141414, "test_eval_f1":
0.22764227642276422, "test_eval_runtime": 1.6884,
"test_eval_samples_per_second": 108.977, "test_eval_steps_per_second": 1.185,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T05:44:38.583882-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.11.rel_attn.q", "transformer.layer.11.rel_attn.k",
"transformer.layer.11.rel_attn.v", "transformer.layer.11.rel_attn.o",
"transformer.layer.11.rel_attn.r", "transformer.layer.11.rel_attn.r_r_bias",
"transformer.layer.11.rel_attn.r_s_bias",
"transformer.layer.11.rel_attn.r_w_bias",
"transformer.layer.11.rel_attn.seg_embed",
"transformer.layer.11.rel_attn.layer_norm.weight",

```



```

"transformer.layer.11.rel_attn.layer_norm.bias",
"transformer.layer.11.ff.layer_norm.weight",
"transformer.layer.11.ff.layer_norm.bias",
"transformer.layer.11.ff.layer_1.weight",
"transformer.layer.11.ff.layer_1.bias",
"transformer.layer.11.ff.layer_2.weight",
"transformer.layer.11.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 117310466, "model_trainable_params":
8270594, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 768, "n_layer": 12, "n_head": 12, "d_head": 64,
"ff_activation": "gelu", "d_inner": 3072, "untie_r": true, "attn_type": "bi",
"initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1, "mem_len":
null, "reuse_len": null, "bi_data": false, "clamp_len": -1, "same_length":
false, "summary_type": "last", "summary_use_proj": true, "summary_activation":
"tanh", "summary_last_dropout": 0.1, "start_n_top": 5, "end_n_top": 5,
"bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2, "use_mems_eval": true,
"use_mems_train": false, "return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-base-cased",
"_commit_hash": "ceaa69c7bc5e512b5007106a7ccbb7daf24b2c79",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}, "run_metrics":
{"train_eval_loss": 0.69793301820755, "train_eval_accuracy": 0.5056031641397495,
"train_eval_precision": 0.515625, "train_eval_recall": 0.1741424802110818,
"train_eval_f1": 0.2603550295857988, "train_eval_runtime": 5.4257,
"train_eval_samples_per_second": 279.593, "train_eval_steps_per_second": 2.212,
"train_epoch": 1.0, "val_eval_loss": 0.7101064324378967, "val_eval_accuracy":
0.5050505050505051, "val_eval_precision": 0.5714285714285714, "val_eval_recall":
0.1568627450980392, "val_eval_f1": 0.24615384615384617, "val_eval_runtime":

```

```

1.545, "val_eval_samples_per_second": 64.079, "val_eval_steps_per_second":
0.647, "val_epoch": 1.0, "test_eval_loss": 0.7075966596603394,
"test_eval_accuracy": 0.4891304347826087, "test_eval_precision":
0.5675675675675675, "test_eval_recall": 0.212121212121213, "test_eval_f1":
0.3088235294117647, "test_eval_runtime": 1.7853, "test_eval_samples_per_second":
103.062, "test_eval_steps_per_second": 1.12, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T05:45:32.331635-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.23.rel_attn.q", "transformer.layer.23.rel_attn.k",
"transformer.layer.23.rel_attn.v", "transformer.layer.23.rel_attn.o",
"transformer.layer.23.rel_attn.r", "transformer.layer.23.rel_attn.r_r_bias",
"transformer.layer.23.rel_attn.r_s_bias",
"transformer.layer.23.rel_attn.r_w_bias",
"transformer.layer.23.rel_attn.seg_embed",
"transformer.layer.23.rel_attn.layer_norm.weight",
"transformer.layer.23.rel_attn.layer_norm.bias",
"transformer.layer.23.ff.layer_norm.weight",
"transformer.layer.23.ff.layer_norm.bias",
"transformer.layer.23.ff.layer_1.weight",
"transformer.layer.23.ff.layer_1.bias",
"transformer.layer.23.ff.layer_2.weight",
"transformer.layer.23.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 361320450, "model_trainable_params":
14697474, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 1024, "n_layer": 24, "n_head": 16, "d_head":
64, "ff_activation": "gelu", "d_inner": 4096, "untie_r": true, "attn_type":
"bi", "initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1,
"mem_len": null, "reuse_len": null, "bi_data": false, "clamp_len": -1,
"same_length": false, "summary_type": "last", "summary_use_proj": true,
"summary_activation": "tanh", "summary_last_dropout": 0.1, "start_n_top": 5,
"end_n_top": 5, "bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2,
"use_mems_eval": true, "use_mems_train": false, "return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,

```

```

"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-large-cased",
"_commit_hash": "37658b4f179eaf971d127d16bfbd9ca676a93034",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}, "run_metrics":
{"train_eval_loss": 0.6783477663993835, "train_eval_accuracy":
0.5754779169413315, "train_eval_precision": 0.5721518987341773,
"train_eval_recall": 0.5963060686015831, "train_eval_f1": 0.5839793281653747,
"train_eval_runtime": 14.4978, "train_eval_samples_per_second": 104.636,
"train_eval_steps_per_second": 0.828, "train_epoch": 1.0, "val_eval_loss":
0.6901865601539612, "val_eval_accuracy": 0.5252525252525253,
"val_eval_precision": 0.532258064516129, "val_eval_recall": 0.6470588235294118,
"val_eval_f1": 0.584070796460177, "val_eval_runtime": 1.9643,
"val_eval_samples_per_second": 50.4, "val_eval_steps_per_second": 0.509,
"val_epoch": 1.0, "test_eval_loss": 0.7030124664306641, "test_eval_accuracy":
0.532608695652174, "test_eval_precision": 0.5656565656565656,
"test_eval_recall": 0.5656565656565656, "test_eval_f1": 0.5656565656565656,
"test_eval_runtime": 2.8135, "test_eval_samples_per_second": 65.399,
"test_eval_steps_per_second": 0.711, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:05:32.220093-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":

```

```

0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdp", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6922009587287903, "train_eval_accuracy": 0.5176194205168363,
"train_eval_precision": 0.5150074294205051, "train_eval_recall":
0.45641295759810374, "train_eval_f1": 0.4839430326724379, "train_eval_runtime":
12.2198, "train_eval_samples_per_second": 627.017,
"train_eval_steps_per_second": 4.91, "train_epoch": 1.0, "val_eval_loss":
0.6900454759597778, "val_eval_accuracy": 0.5201900237529691,
"val_eval_precision": 0.4731182795698925, "val_eval_recall": 0.4583333333333333,
"val_eval_f1": 0.4656084656084656, "val_eval_runtime": 1.6523,
"val_eval_samples_per_second": 254.795, "val_eval_steps_per_second": 2.421,
"val_epoch": 1.0, "test_eval_loss": 0.6946691870689392, "test_eval_accuracy":
0.49073064340239914, "test_eval_precision": 0.4711111111111111,
"test_eval_recall": 0.48072562358276644, "test_eval_f1": 0.47586980920314254,
"test_eval_runtime": 3.1947, "test_eval_samples_per_second": 287.035,
"test_eval_steps_per_second": 2.504, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:06:35.344523-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_pos_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":

```

```

"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.695508599281311, "train_eval_accuracy": 0.5044374836857217,
"train_eval_precision": 0.0, "train_eval_recall": 0.0, "train_eval_f1": 0.0,
"train_eval_runtime": 12.2492, "train_eval_samples_per_second": 625.512,
"train_eval_steps_per_second": 4.898, "train_epoch": 1.0, "val_eval_loss":
0.6879574060440063, "val_eval_accuracy": 0.5439429928741093,
"val_eval_precision": 0.0, "val_eval_recall": 0.0, "val_eval_f1": 0.0,
"val_eval_runtime": 1.9018, "val_eval_samples_per_second": 221.37,
"val_eval_steps_per_second": 2.103, "val_epoch": 1.0, "test_eval_loss":
0.6950060725212097, "test_eval_accuracy": 0.5190839694656488,
"test_eval_precision": 0.0, "test_eval_recall": 0.0, "test_eval_f1": 0.0,
"test_eval_runtime": 2.3847, "test_eval_samples_per_second": 384.529,
"test_eval_steps_per_second": 3.355, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:07:45.164432-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":

```

```

"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6928959488868713, "train_eval_accuracy": 0.509397024275646,
"train_eval_precision": 0.5033089515848137, "train_eval_recall":
0.7611272056887016, "train_eval_f1": 0.6059335360100639, "train_eval_runtime":
12.2449, "train_eval_samples_per_second": 625.728,
"train_eval_steps_per_second": 4.9, "train_epoch": 1.0, "val_eval_loss":
0.6924623847007751, "val_eval_accuracy": 0.5083135391923991,
"val_eval_precision": 0.47648902821316613, "val_eval_recall":
0.7916666666666666, "val_eval_f1": 0.5949119373776908, "val_eval_runtime":
1.6899, "val_eval_samples_per_second": 249.129, "val_eval_steps_per_second":
2.367, "val_epoch": 1.0, "test_eval_loss": 0.6922692656517029,
"test_eval_accuracy": 0.5136314067611778, "test_eval_precision":
0.4964936886395512, "test_eval_recall": 0.8027210884353742, "test_eval_f1":
0.6135181975736569, "test_eval_runtime": 2.6667, "test_eval_samples_per_second":
343.868, "test_eval_steps_per_second": 3.0, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:08:29.015206-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",

```

```

"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6929386854171753,
"train_eval_accuracy": 0.5194466196815453, "train_eval_precision":
0.5097375105842507, "train_eval_recall": 0.7927311035027653, "train_eval_f1":
0.6204906204906205, "train_eval_runtime": 12.7522,
"train_eval_samples_per_second": 600.836, "train_eval_steps_per_second": 4.705,
"train_epoch": 1.0, "val_eval_loss": 0.6928949952125549, "val_eval_accuracy":
0.505938242280285, "val_eval_precision": 0.4751552795031056, "val_eval_recall":
0.796875, "val_eval_f1": 0.5953307392996109, "val_eval_runtime": 1.7282,
"val_eval_samples_per_second": 243.612, "val_eval_steps_per_second": 2.315,
"val_epoch": 1.0, "test_eval_loss": 0.6930844783782959, "test_eval_accuracy":
0.5005452562704471, "test_eval_precision": 0.48787446504992865,
"test_eval_recall": 0.7755102040816326, "test_eval_f1": 0.5989492119089317,
"test_eval_runtime": 2.4264, "test_eval_samples_per_second": 377.932,
"test_eval_steps_per_second": 3.297, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:11:12.445054-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,

```

```

"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."], "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaeff65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.7662244439125061,
"train_eval_accuracy": 0.5182719916470895, "train_eval_precision":
0.5202754399387911, "train_eval_recall": 0.358177508559389, "train_eval_f1":
0.4242707845889877, "train_eval_runtime": 35.7971,
"train_eval_samples_per_second": 214.04, "train_eval_steps_per_second": 1.676,
"train_epoch": 1.0, "val_eval_loss": 0.7611860036849976, "val_eval_accuracy":
0.5249406175771971, "val_eval_precision": 0.47368421052631576,
"val_eval_recall": 0.375, "val_eval_f1": 0.4186046511627907, "val_eval_runtime":
3.7811, "val_eval_samples_per_second": 111.342, "val_eval_steps_per_second":
1.058, "val_epoch": 1.0, "test_eval_loss": 0.768562912940979,
"test_eval_accuracy": 0.5212649945474372, "test_eval_precision":
0.5031847133757962, "test_eval_recall": 0.35827664399092973, "test_eval_f1":

```



```

0.4185430463576159, "test_eval_runtime": 5.3123, "test_eval_samples_per_second":
172.618, "test_eval_steps_per_second": 1.506, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:13:01.437684-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,

```

```

"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6928300261497498,
"train_eval_accuracy": 0.49960845732184805, "train_eval_precision":
0.49728220949023066, "train_eval_recall": 0.8914932841717145, "train_eval_f1":
0.6384383251603168, "train_eval_runtime": 35.9761,
"train_eval_samples_per_second": 212.975, "train_eval_steps_per_second": 1.668,
"train_epoch": 1.0, "val_eval_loss": 0.6941162347793579, "val_eval_accuracy":
0.45605700712589076, "val_eval_precision": 0.45066666666666666,
"val_eval_recall": 0.8802083333333334, "val_eval_f1": 0.5961199294532628,
"val_eval_runtime": 2.9929, "val_eval_samples_per_second": 140.668,
"val_eval_steps_per_second": 1.337, "val_epoch": 1.0, "test_eval_loss":
0.6934722065925598, "test_eval_accuracy": 0.48964013086150493,
"test_eval_precision": 0.4833538840937115, "test_eval_recall":
0.8888888888888888, "test_eval_f1": 0.6261980830670927, "test_eval_runtime":
5.3144, "test_eval_samples_per_second": 172.552, "test_eval_steps_per_second":
1.505, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:14:16.909961-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},

```

```

"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.9958739876747131, "train_eval_accuracy":
0.49334377447141736, "train_eval_precision": 0.49433861729052886,
"train_eval_recall": 0.977350539899921, "train_eval_f1": 0.6565817409766455,
"train_eval_runtime": 16.6352, "train_eval_samples_per_second": 460.589,
"train_eval_steps_per_second": 3.607, "train_epoch": 1.0, "val_eval_loss":
1.0358781814575195, "val_eval_accuracy": 0.4679334916864608,
"val_eval_precision": 0.46078431372549017, "val_eval_recall":
0.9791666666666666, "val_eval_f1": 0.6266666666666667, "val_eval_runtime":
2.0038, "val_eval_samples_per_second": 210.105, "val_eval_steps_per_second":
1.996, "val_epoch": 1.0, "test_eval_loss": 1.0367566347122192,
"test_eval_accuracy": 0.4830970556161396, "test_eval_precision":
0.4816870144284129, "test_eval_recall": 0.9841269841269841, "test_eval_f1":
0.646795827123696, "test_eval_runtime": 3.0183, "test_eval_samples_per_second":
303.811, "test_eval_steps_per_second": 2.65, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:16:30.328575-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-large", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.27.attn_norm.weight", "model.layers.27.attn.Wqkv.weight",
"model.layers.27.attn.Wo.weight", "model.layers.27.mlp_norm.weight",
"model.layers.27.mlp.Wi.weight", "model.layers.27.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 395833346, "model_trainable_params": 13309954,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,

```

```

"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-large",
"_commit_hash": "45bb4654a4d5aaff24dd11d4781fa46d39bf8c13",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 1024, "intermediate_size": 2624,
"num_hidden_layers": 28, "num_attention_heads": 16, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rop_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rop_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.6922593116760254, "train_eval_accuracy":
0.5356303837118246, "train_eval_precision": 0.5326235326235326,
"train_eval_recall": 0.5138267052936529, "train_eval_f1": 0.5230563002680965,
"train_eval_runtime": 43.4543, "train_eval_samples_per_second": 176.323,
"train_eval_steps_per_second": 1.381, "train_epoch": 1.0, "val_eval_loss":
0.6889220476150513, "val_eval_accuracy": 0.5581947743467933,
"val_eval_precision": 0.5157894736842106, "val_eval_recall": 0.5104166666666666,
"val_eval_f1": 0.5130890052356021, "val_eval_runtime": 4.4502,
"val_eval_samples_per_second": 94.602, "val_eval_steps_per_second": 0.899,
"val_epoch": 1.0, "test_eval_loss": 0.6935679912567139, "test_eval_accuracy":
0.5332606324972737, "test_eval_precision": 0.5147392290249433,
"test_eval_recall": 0.5147392290249433, "test_eval_f1": 0.5147392290249433,
"test_eval_runtime": 6.1901, "test_eval_samples_per_second": 148.139,
"test_eval_steps_per_second": 1.292, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:17:40.883299-07:00", "experiment_meta":

```

```

{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",
"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",
"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":
"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":
null, "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":
"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":

```

```

12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,
"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.6928727030754089, "train_eval_accuracy":
0.5077003393369878, "train_eval_precision": 0.5046313449425713,
"train_eval_recall": 0.3587042401896234, "train_eval_f1": 0.41933497536945813,
"train_eval_runtime": 15.9213, "train_eval_samples_per_second": 481.242,
"train_eval_steps_per_second": 3.769, "train_epoch": 1.0, "val_eval_loss":
0.6966883540153503, "val_eval_accuracy": 0.47743467933491684,
"val_eval_precision": 0.41025641025641024, "val_eval_recall":
0.3333333333333333, "val_eval_f1": 0.367816091954023, "val_eval_runtime":
2.0795, "val_eval_samples_per_second": 202.457, "val_eval_steps_per_second":
1.924, "val_epoch": 1.0, "test_eval_loss": 0.6913263201713562,
"test_eval_accuracy": 0.5223555070883316, "test_eval_precision":
0.504225352112676, "test_eval_recall": 0.40589569160997735, "test_eval_f1":
0.44974874371859297, "test_eval_runtime": 3.203, "test_eval_samples_per_second":
286.295, "test_eval_steps_per_second": 2.498, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:20:05.678048-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.11.rel_attn.q", "transformer.layer.11.rel_attn.k",
"transformer.layer.11.rel_attn.v", "transformer.layer.11.rel_attn.o",
"transformer.layer.11.rel_attn.r", "transformer.layer.11.rel_attn.r_r_bias",
"transformer.layer.11.rel_attn.r_s_bias",
"transformer.layer.11.rel_attn.r_w_bias",
"transformer.layer.11.rel_attn.seg_embed",
"transformer.layer.11.rel_attn.layer_norm.weight",
"transformer.layer.11.rel_attn.layer_norm.bias",
"transformer.layer.11.ff.layer_norm.weight",
"transformer.layer.11.ff.layer_norm.bias",
"transformer.layer.11.ff.layer_1.weight",
"transformer.layer.11.ff.layer_1.bias",
"transformer.layer.11.ff.layer_2.weight",
"transformer.layer.11.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 117310466, "model_trainable_params":
8270594, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 768, "n_layer": 12, "n_head": 12, "d_head": 64,
"ff_activation": "gelu", "d_inner": 3072, "untie_r": true, "attn_type": "bi",
"initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1, "mem_len":
null, "reuse_len": null, "bi_data": false, "clamp_len": -1, "same_length":
false, "summary_type": "last", "summary_use_proj": true, "summary_activation":
"tanh", "summary_last_dropout": 0.1, "start_n_top": 5, "end_n_top": 5,

```

```

"bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2, "use_mems_eval": true,
"use_mems_train": false, "return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-base-cased",
"_commit_hash": "ceaa69c7bc5e512b5007106a7ccbb7daf24b2c79",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}, "run_metrics":
{"train_eval_loss": 0.6961987614631653, "train_eval_accuracy":
0.4976507439310885, "train_eval_precision": 0.49203919167176974,
"train_eval_recall": 0.42322886489333683, "train_eval_f1": 0.4550474302704233,
"train_eval_runtime": 20.0405, "train_eval_samples_per_second": 382.326,
"train_eval_steps_per_second": 2.994, "train_epoch": 1.0, "val_eval_loss":
0.685169517993927, "val_eval_accuracy": 0.5558194774346793,
"val_eval_precision": 0.5133689839572193, "val_eval_recall": 0.5, "val_eval_f1":
0.5065963060686016, "val_eval_runtime": 2.2786, "val_eval_samples_per_second":
184.766, "val_eval_steps_per_second": 1.756, "val_epoch": 1.0, "test_eval_loss":
0.6974068880081177, "test_eval_accuracy": 0.48636859323882226,
"test_eval_precision": 0.4623115577889447, "test_eval_recall":
0.41723356009070295, "test_eval_f1": 0.43861740166865315, "test_eval_runtime":
3.4268, "test_eval_samples_per_second": 267.594, "test_eval_steps_per_second":
2.335, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:22:43.158709-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.23.rel_attn.q", "transformer.layer.23.rel_attn.k",
"transformer.layer.23.rel_attn.v", "transformer.layer.23.rel_attn.o",
"transformer.layer.23.rel_attn.r", "transformer.layer.23.rel_attn.r_r_bias",
"transformer.layer.23.rel_attn.r_s_bias",
"transformer.layer.23.rel_attn.r_w_bias",

```

```

"transformer.layer.23.rel_attn.seg_embed",
"transformer.layer.23.rel_attn.layer_norm.weight",
"transformer.layer.23.rel_attn.layer_norm.bias",
"transformer.layer.23.ff.layer_norm.weight",
"transformer.layer.23.ff.layer_norm.bias",
"transformer.layer.23.ff.layer_1.weight",
"transformer.layer.23.ff.layer_1.bias",
"transformer.layer.23.ff.layer_2.weight",
"transformer.layer.23.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 361320450, "model_trainable_params":
14697474, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 1024, "n_layer": 24, "n_head": 16, "d_head":
64, "ff_activation": "gelu", "d_inner": 4096, "untie_r": true, "attn_type":
"bi", "initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1,
"mem_len": null, "reuse_len": null, "bi_data": false, "clamp_len": -1,
"same_length": false, "summary_type": "last", "summary_use_proj": true,
"summary_activation": "tanh", "summary_last_dropout": 0.1, "start_n_top": 5,
"end_n_top": 5, "bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2,
"use_mems_eval": true, "use_mems_train": false, "return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-large-cased",
"_commit_hash": "37658b4f179eaf971d127d16bfbd9ca676a93034",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}}, "run_metrics":
{"train_eval_loss": 0.6941739916801453, "train_eval_accuracy":
0.5211433046202036, "train_eval_precision": 0.515625, "train_eval_recall":
0.5562286015275217, "train_eval_f1": 0.5351577347016344, "train_eval_runtime":
54.1542, "train_eval_samples_per_second": 141.485,
"train_eval_steps_per_second": 1.108, "train_epoch": 1.0, "val_eval_loss":

```



```

0.7050549983978271, "val_eval_accuracy": 0.47980997624703087,
"val_eval_precision": 0.43891402714932126, "val_eval_recall":
0.5052083333333334, "val_eval_f1": 0.46973365617433416, "val_eval_runtime":
3.9951, "val_eval_samples_per_second": 105.38, "val_eval_steps_per_second":
1.001, "val_epoch": 1.0, "test_eval_loss": 0.702609121799469,
"test_eval_accuracy": 0.49291166848418755, "test_eval_precision":
0.4745762711864407, "test_eval_recall": 0.5079365079365079, "test_eval_f1":
0.49069003285870755, "test_eval_runtime": 7.4402,
"test_eval_samples_per_second": 123.25, "test_eval_steps_per_second": 1.075,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:26:26.369648-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_dep_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6945847272872925, "train_eval_accuracy": 0.5165753067084312,
"train_eval_precision": 0.5692995529061102, "train_eval_recall":

```

```

0.10060574137476956, "train_eval_f1": 0.170993733213966, "train_eval_runtime":
12.5461, "train_eval_samples_per_second": 610.709,
"train_eval_steps_per_second": 4.782, "train_epoch": 1.0, "val_eval_loss":
0.6942901015281677, "val_eval_accuracy": 0.5415676959619953,
"val_eval_precision": 0.4864864864864865, "val_eval_recall": 0.09375,
"val_eval_f1": 0.1572052401746725, "val_eval_runtime": 1.6397,
"val_eval_samples_per_second": 256.754, "val_eval_steps_per_second": 2.439,
"val_epoch": 1.0, "test_eval_loss": 0.6943947672843933, "test_eval_accuracy":
0.5169029443838604, "test_eval_precision": 0.4891304347826087,
"test_eval_recall": 0.10204081632653061, "test_eval_f1": 0.16885553470919323,
"test_eval_runtime": 2.3631, "test_eval_samples_per_second": 388.052,
"test_eval_steps_per_second": 3.385, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:27:25.366089-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":

```

```

0.6943996548652649, "train_eval_accuracy": 0.5086139389193423,
"train_eval_precision": 0.5098765432098765, "train_eval_recall":
0.21754016328680537, "train_eval_f1": 0.30496584825549194, "train_eval_runtime":
16.2442, "train_eval_samples_per_second": 471.676,
"train_eval_steps_per_second": 3.694, "train_epoch": 1.0, "val_eval_loss":
0.6945574879646301, "val_eval_accuracy": 0.5106888361045131,
"val_eval_precision": 0.4270833333333333, "val_eval_recall":
0.21354166666666666, "val_eval_f1": 0.2847222222222222, "val_eval_runtime":
5.4366, "val_eval_samples_per_second": 77.438, "val_eval_steps_per_second":
0.736, "val_epoch": 1.0, "test_eval_loss": 0.6916611790657043,
"test_eval_accuracy": 0.5256270447110142, "test_eval_precision":
0.5164835164835165, "test_eval_recall": 0.21315192743764172, "test_eval_f1":
0.3017656500802568, "test_eval_runtime": 6.1429, "test_eval_samples_per_second":
149.279, "test_eval_steps_per_second": 1.302, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:28:37.121108-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,

```

```

"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6932464241981506,
"train_eval_accuracy": 0.4955625163142783, "train_eval_precision":
0.4948624962224237, "train_eval_recall": 0.8625230445088228, "train_eval_f1":
0.628900624099856, "train_eval_runtime": 16.1, "train_eval_samples_per_second":
475.9, "train_eval_steps_per_second": 3.727, "train_epoch": 1.0,
"val_eval_loss": 0.6933861970901489, "val_eval_accuracy": 0.45605700712589076,
"val_eval_precision": 0.44931506849315067, "val_eval_recall":
0.8541666666666666, "val_eval_f1": 0.5888689407540395, "val_eval_runtime":
5.4728, "val_eval_samples_per_second": 76.926, "val_eval_steps_per_second":
0.731, "val_epoch": 1.0, "test_eval_loss": 0.6933544874191284,
"test_eval_accuracy": 0.48091603053435117, "test_eval_precision":
0.47782002534854245, "test_eval_recall": 0.854875283446712, "test_eval_f1":
0.6130081300813008, "test_eval_runtime": 6.1754, "test_eval_samples_per_second":
148.493, "test_eval_steps_per_second": 1.295, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:29:58.973720-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",

```

```

"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6931515336036682,
"train_eval_accuracy": 0.4984338292873923, "train_eval_precision":
0.4968613537117904, "train_eval_recall": 0.9589149328417171, "train_eval_f1":
0.654561797752809, "train_eval_runtime": 16.2685,
"train_eval_samples_per_second": 470.973, "train_eval_steps_per_second": 3.688,
"train_epoch": 1.0, "val_eval_loss": 0.6935935020446777, "val_eval_accuracy":
0.44655581947743467, "val_eval_precision": 0.4481012658227848,
"val_eval_recall": 0.921875, "val_eval_f1": 0.6030664395229983,
"val_eval_runtime": 5.6563, "val_eval_samples_per_second": 74.431,
"val_eval_steps_per_second": 0.707, "val_epoch": 1.0, "test_eval_loss":
0.6933175325393677, "test_eval_accuracy": 0.4820065430752454,
"test_eval_precision": 0.480681818181815, "test_eval_recall":

```

```

0.9591836734693877, "test_eval_f1": 0.6404239212717638, "test_eval_runtime":
6.1416, "test_eval_samples_per_second": 149.31, "test_eval_steps_per_second":
1.303, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:33:28.334852-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaef65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6947011947631836,
"train_eval_accuracy": 0.5005220569042026, "train_eval_precision":
0.49451754385964913, "train_eval_recall": 0.3563339478535686, "train_eval_f1":
0.4142048063676718, "train_eval_runtime": 39.2905,
"train_eval_samples_per_second": 195.009, "train_eval_steps_per_second": 1.527,
"train_epoch": 1.0, "val_eval_loss": 0.6937609314918518, "val_eval_accuracy":
0.48931116389548696, "val_eval_precision": 0.4148148148148148,

```

```

"val_eval_recall": 0.2916666666666667, "val_eval_f1": 0.3425076452599388,
"val_eval_runtime": 7.5497, "val_eval_samples_per_second": 55.764,
"val_eval_steps_per_second": 0.53, "val_epoch": 1.0, "test_eval_loss":
0.6913156509399414, "test_eval_accuracy": 0.52453653217012,
"test_eval_precision": 0.5077399380804953, "test_eval_recall":
0.37188208616780044, "test_eval_f1": 0.4293193717277487, "test_eval_runtime":
9.1882, "test_eval_samples_per_second": 99.802, "test_eval_steps_per_second":
0.871, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:35:44.343823-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},

```

```

"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdp", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6936127543449402,
"train_eval_accuracy": 0.49712868702688595, "train_eval_precision":
0.49601706970128023, "train_eval_recall": 0.9183565973136687, "train_eval_f1":
0.6441304147039808, "train_eval_runtime": 39.7335,
"train_eval_samples_per_second": 192.835, "train_eval_steps_per_second": 1.51,
"train_epoch": 1.0, "val_eval_loss": 0.6945639252662659, "val_eval_accuracy":
0.4750593824228028, "val_eval_precision": 0.46153846153846156,
"val_eval_recall": 0.90625, "val_eval_f1": 0.6115992970123023,
"val_eval_runtime": 6.8431, "val_eval_samples_per_second": 61.522,
"val_eval_steps_per_second": 0.585, "val_epoch": 1.0, "test_eval_loss":
0.6940426826477051, "test_eval_accuracy": 0.49073064340239914,
"test_eval_precision": 0.484375, "test_eval_recall": 0.9138321995464853,
"test_eval_f1": 0.6331500392772977, "test_eval_runtime": 9.9284,
"test_eval_samples_per_second": 92.361, "test_eval_steps_per_second": 0.806,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:37:22.800464-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,

```



```

"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.7583022713661194, "train_eval_accuracy":
0.495301487862177, "train_eval_precision": 0.49308300395256915,
"train_eval_recall": 0.6570977087174085, "train_eval_f1": 0.5633961838094163,
"train_eval_runtime": 20.2852, "train_eval_samples_per_second": 377.713,
"train_eval_steps_per_second": 2.958, "train_epoch": 1.0, "val_eval_loss":
0.7390754222869873, "val_eval_accuracy": 0.5273159144893111,
"val_eval_precision": 0.48638132295719844, "val_eval_recall":
0.6510416666666666, "val_eval_f1": 0.5567928730512249, "val_eval_runtime":
5.6727, "val_eval_samples_per_second": 74.216, "val_eval_steps_per_second":
0.705, "val_epoch": 1.0, "test_eval_loss": 0.7513620257377625,
"test_eval_accuracy": 0.49291166848418755, "test_eval_precision":
0.47923875432525953, "test_eval_recall": 0.6281179138321995, "test_eval_f1":
0.5436702649656526, "test_eval_runtime": 6.6508, "test_eval_samples_per_second":
137.878, "test_eval_steps_per_second": 1.203, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:39:58.067178-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-large", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.27.attn_norm.weight", "model.layers.27.attn.Wqkv.weight",
"model.layers.27.attn.Wo.weight", "model.layers.27.mlp_norm.weight",
"model.layers.27.mlp.Wi.weight", "model.layers.27.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",

```

```

"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 395833346, "model_trainable_params": 13309954,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-large",
"_commit_hash": "45bb4654a4d5aaff24dd11d4781fa46d39bf8c13",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 1024, "intermediate_size": 2624,
"num_hidden_layers": 28, "num_attention_heads": 16, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.6982264518737793, "train_eval_accuracy":
0.5157922213521274, "train_eval_precision": 0.5101045296167247,
"train_eval_recall": 0.5783513299973664, "train_eval_f1": 0.542088373241175,
"train_eval_runtime": 46.3094, "train_eval_samples_per_second": 165.452,
"train_eval_steps_per_second": 1.296, "train_epoch": 1.0, "val_eval_loss":
0.6816452145576477, "val_eval_accuracy": 0.5629453681710214,
"val_eval_precision": 0.5185185185185185, "val_eval_recall": 0.5833333333333334,
"val_eval_f1": 0.5490196078431373, "val_eval_runtime": 7.0301,
"val_eval_samples_per_second": 59.885, "val_eval_steps_per_second": 0.569,

```

```

"val_epoch": 1.0, "test_eval_loss": 0.6970695853233337, "test_eval_accuracy":
0.5125408942202835, "test_eval_precision": 0.49427480916030536,
"test_eval_recall": 0.5873015873015873, "test_eval_f1": 0.5367875647668394,
"test_eval_runtime": 10.6884, "test_eval_samples_per_second": 85.794,
"test_eval_steps_per_second": 0.748, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:41:32.634630-07:00", "experiment_meta":
{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",
"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",
"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,

```

```

"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":
"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":
null, "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":
"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":
12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,
"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.6943697333335876, "train_eval_accuracy":
0.5088749673714434, "train_eval_precision": 0.5039553280595626,
"train_eval_recall": 0.5704503555438504, "train_eval_f1": 0.5351451513279802,
"train_eval_runtime": 19.2162, "train_eval_samples_per_second": 398.726,
"train_eval_steps_per_second": 3.122, "train_epoch": 1.0, "val_eval_loss":
0.6942377686500549, "val_eval_accuracy": 0.49406175771971494,
"val_eval_precision": 0.45414847161572053, "val_eval_recall":
0.5416666666666666, "val_eval_f1": 0.49406175771971494, "val_eval_runtime":
5.6593, "val_eval_samples_per_second": 74.391, "val_eval_steps_per_second":
0.707, "val_epoch": 1.0, "test_eval_loss": 0.6941428184509277,
"test_eval_accuracy": 0.495092693565976, "test_eval_precision":
0.4791666666666667, "test_eval_recall": 0.5736961451247166, "test_eval_f1":
0.52218782249742, "test_eval_runtime": 6.7597, "test_eval_samples_per_second":
135.657, "test_eval_steps_per_second": 1.183, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:44:18.434030-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.11.rel_attn.q", "transformer.layer.11.rel_attn.k",
"transformer.layer.11.rel_attn.v", "transformer.layer.11.rel_attn.o",
"transformer.layer.11.rel_attn.r", "transformer.layer.11.rel_attn.r_r_bias",
"transformer.layer.11.rel_attn.r_s_bias",
"transformer.layer.11.rel_attn.r_w_bias",
"transformer.layer.11.rel_attn.seg_embed",
"transformer.layer.11.rel_attn.layer_norm.weight",
"transformer.layer.11.rel_attn.layer_norm.bias",
"transformer.layer.11.ff.layer_norm.weight",
"transformer.layer.11.ff.layer_norm.bias",
"transformer.layer.11.ff.layer_1.weight",
"transformer.layer.11.ff.layer_1.bias",
"transformer.layer.11.ff.layer_2.weight",
"transformer.layer.11.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 117310466, "model_trainable_params":
8270594, "pooler_activation_function": "N/A", "config_attributes":

```

```

{"vocab_size": 32000, "d_model": 768, "n_layer": 12, "n_head": 12, "d_head": 64,
"ff_activation": "gelu", "d_inner": 3072, "untie_r": true, "attn_type": "bi",
"initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1, "mem_len":
null, "reuse_len": null, "bi_data": false, "clamp_len": -1, "same_length":
false, "summary_type": "last", "summary_use_proj": true, "summary_activation":
"tanh", "summary_last_dropout": 0.1, "start_n_top": 5, "end_n_top": 5,
"bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2, "use_mems_eval": true,
"use_mems_train": false, "return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-base-cased",
"_commit_hash": "ceaa69c7bc5e512b5007106a7ccbb7daf24b2c79",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}, {"run_metrics":
{"train_eval_loss": 0.6962189674377441, "train_eval_accuracy":
0.49752022970503784, "train_eval_precision": 0.4916482823826032,
"train_eval_recall": 0.41085067158282856, "train_eval_f1": 0.44763271162123386,
"train_eval_runtime": 23.4281, "train_eval_samples_per_second": 327.043,
"train_eval_steps_per_second": 2.561, "train_epoch": 1.0, "val_eval_loss":
0.6976354718208313, "val_eval_accuracy": 0.5106888361045131,
"val_eval_precision": 0.46354166666666667, "val_eval_recall": 0.46354166666666667,
"val_eval_f1": 0.46354166666666667, "val_eval_runtime": 5.8562,
"val_eval_samples_per_second": 71.889, "val_eval_steps_per_second": 0.683,
"val_epoch": 1.0, "test_eval_loss": 0.6960442662239075, "test_eval_accuracy":
0.4830970556161396, "test_eval_precision": 0.45905707196029777,
"test_eval_recall": 0.41950113378684806, "test_eval_f1": 0.43838862559241704,
"test_eval_runtime": 6.9733, "test_eval_samples_per_second": 131.502,
"test_eval_steps_per_second": 1.147, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:46:51.772831-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":

```

```

"snc_morph_complexity_value", "y_col": "binary_complexity",
"layers_to_unfreeze": ["bert.encoder.layer.11.", "bert.pooler.",
"classifier."], "model_details": {"model_total_params": 108311810,
"model_trainable_params": 7680002, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6955558061599731, "train_eval_accuracy": 0.4939963456016706,
"train_eval_precision": 0.49298245614035086, "train_eval_recall":
0.7400579404793258, "train_eval_f1": 0.5917658207855112, "train_eval_runtime":
16.2252, "train_eval_samples_per_second": 472.229,
"train_eval_steps_per_second": 3.698, "train_epoch": 1.0, "val_eval_loss":
0.6970760226249695, "val_eval_accuracy": 0.46318289786223277,
"val_eval_precision": 0.4472049689440994, "val_eval_recall": 0.75,
"val_eval_f1": 0.5603112840466926, "val_eval_runtime": 5.3951,
"val_eval_samples_per_second": 78.034, "val_eval_steps_per_second": 0.741,
"val_epoch": 1.0, "test_eval_loss": 0.6958844065666199, "test_eval_accuracy":
0.4765539803707743, "test_eval_precision": 0.47040971168437024,
"test_eval_recall": 0.7029478458049887, "test_eval_f1": 0.5636363636363636,
"test_eval_runtime": 7.4313, "test_eval_samples_per_second": 123.397,
"test_eval_steps_per_second": 1.077, "test_epoch": 1.0}}

```

```

{"timestamp": "2025-04-11T08:48:04.426445-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_complexity_value", "y_col": "binary_complexity",
"layers_to_unfreeze": ["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",

```

```

"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6931843757629395,
"train_eval_accuracy": 0.5037849125554685, "train_eval_precision":
0.49964402676918696, "train_eval_recall": 0.924150645246247, "train_eval_f1":
0.6486136783733826, "train_eval_runtime": 16.1503,
"train_eval_samples_per_second": 474.417, "train_eval_steps_per_second": 3.715,
"train_epoch": 1.0, "val_eval_loss": 0.6934304237365723, "val_eval_accuracy":
0.47980997624703087, "val_eval_precision": 0.46582278481012657,
"val_eval_recall": 0.9583333333333334, "val_eval_f1": 0.626916524701874,
"val_eval_runtime": 5.5163, "val_eval_samples_per_second": 76.319,
"val_eval_steps_per_second": 0.725, "val_epoch": 1.0, "test_eval_loss":
0.6931877732276917, "test_eval_accuracy": 0.495092693565976,
"test_eval_precision": 0.486810551558753, "test_eval_recall":
0.9206349206349206, "test_eval_f1": 0.6368627450980392, "test_eval_runtime":
7.3098, "test_eval_samples_per_second": 125.448, "test_eval_steps_per_second":
1.094, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:49:30.434362-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_pos_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":

```



```

"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6862639784812927, "train_eval_accuracy": 0.5748187211601846,
"train_eval_precision": 0.5736636245110821, "train_eval_recall":
0.5804749340369393, "train_eval_f1": 0.5770491803278689, "train_eval_runtime":
3.5592, "train_eval_samples_per_second": 426.224, "train_eval_steps_per_second":
3.372, "train_epoch": 1.0, "val_eval_loss": 0.6918572187423706,
"val_eval_accuracy": 0.494949494949495, "val_eval_precision":
0.5087719298245614, "val_eval_recall": 0.5686274509803921, "val_eval_f1":
0.5370370370370371, "val_eval_runtime": 1.1533, "val_eval_samples_per_second":
85.844, "val_eval_steps_per_second": 0.867, "val_epoch": 1.0, "test_eval_loss":
0.6855954527854919, "test_eval_accuracy": 0.5706521739130435,
"test_eval_precision": 0.6063829787234043, "test_eval_recall":
0.5757575757575758, "test_eval_f1": 0.5906735751295337, "test_eval_runtime":
1.2622, "test_eval_samples_per_second": 145.775, "test_eval_steps_per_second":
1.585, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:49:51.656184-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_pos_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":

```

```

false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbacc7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6935935616493225,
"train_eval_accuracy": 0.4996704021094265, "train_eval_precision":
0.4996704021094265, "train_eval_recall": 1.0, "train_eval_f1":
0.6663736263736264, "train_eval_runtime": 3.5845,
"train_eval_samples_per_second": 423.206, "train_eval_steps_per_second": 3.348,
"train_epoch": 1.0, "val_eval_loss": 0.6924985647201538, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 1.6772,
"val_eval_samples_per_second": 59.028, "val_eval_steps_per_second": 0.596,
"val_epoch": 1.0, "test_eval_loss": 0.6921464800834656, "test_eval_accuracy":
0.5380434782608695, "test_eval_precision": 0.5380434782608695,
"test_eval_recall": 1.0, "test_eval_f1": 0.6996466431095406,
"test_eval_runtime": 2.1198, "test_eval_samples_per_second": 86.802,
"test_eval_steps_per_second": 0.944, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:52:34.603555-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_pos_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,

```

```

"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6954520344734192, "train_eval_accuracy": 0.4693473961766645,
"train_eval_precision": 0.47968885047536736, "train_eval_recall":
0.7321899736147758, "train_eval_f1": 0.5796344647519582, "train_eval_runtime":
3.3889, "train_eval_samples_per_second": 447.642, "train_eval_steps_per_second":
3.541, "train_epoch": 1.0, "val_eval_loss": 0.6960757970809937,
"val_eval_accuracy": 0.45454545454545453, "val_eval_precision":
0.4794520547945205, "val_eval_recall": 0.6862745098039216, "val_eval_f1":
0.5645161290322581, "val_eval_runtime": 1.274, "val_eval_samples_per_second":
77.709, "val_eval_steps_per_second": 0.785, "val_epoch": 1.0, "test_eval_loss":
0.6950387358665466, "test_eval_accuracy": 0.4891304347826087,
"test_eval_precision": 0.5174825174825175, "test_eval_recall":
0.7474747474747475, "test_eval_f1": 0.6115702479338843, "test_eval_runtime":
1.5421, "test_eval_samples_per_second": 119.32, "test_eval_steps_per_second":
1.297, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:52:51.650888-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_pos_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",

```

```

"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6935935616493225,
"train_eval_accuracy": 0.4996704021094265, "train_eval_precision":
0.4996704021094265, "train_eval_recall": 1.0, "train_eval_f1":

```

```

0.6663736263736264, "train_eval_runtime": 3.4105,
"train_eval_samples_per_second": 444.803, "train_eval_steps_per_second": 3.519,
"train_epoch": 1.0, "val_eval_loss": 0.6924985647201538, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 1.2536,
"val_eval_samples_per_second": 78.971, "val_eval_steps_per_second": 0.798,
"val_epoch": 1.0, "test_eval_loss": 0.6921464800834656, "test_eval_accuracy":
0.5380434782608695, "test_eval_precision": 0.5380434782608695,
"test_eval_recall": 1.0, "test_eval_f1": 0.6996466431095406,
"test_eval_runtime": 1.3723, "test_eval_samples_per_second": 134.08,
"test_eval_steps_per_second": 1.457, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:54:22.154631-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_pos_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaeff65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,

```

```

"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.7105613350868225,
"train_eval_accuracy": 0.4917600527356625, "train_eval_precision":
0.47796610169491527, "train_eval_recall": 0.18601583113456466, "train_eval_f1":
0.2678062678062678, "train_eval_runtime": 8.2544,
"train_eval_samples_per_second": 183.781, "train_eval_steps_per_second": 1.454,
"train_epoch": 1.0, "val_eval_loss": 0.7215713858604431, "val_eval_accuracy":
0.45454545454545453, "val_eval_precision": 0.43478260869565216,
"val_eval_recall": 0.19607843137254902, "val_eval_f1": 0.2702702702702703,
"val_eval_runtime": 2.0912, "val_eval_samples_per_second": 47.342,
"val_eval_steps_per_second": 0.478, "val_epoch": 1.0, "test_eval_loss":
0.7104301452636719, "test_eval_accuracy": 0.5163043478260869,
"test_eval_precision": 0.6785714285714286, "test_eval_recall":
0.1919191919191919, "test_eval_f1": 0.2992125984251969, "test_eval_runtime":
2.4055, "test_eval_samples_per_second": 76.493, "test_eval_steps_per_second":
0.831, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:55:06.218017-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_pos_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":

```

```

1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6934924125671387,
"train_eval_accuracy": 0.4996704021094265, "train_eval_precision":
0.4996704021094265, "train_eval_recall": 1.0, "train_eval_f1":
0.6663736263736264, "train_eval_runtime": 9.3372,
"train_eval_samples_per_second": 162.468, "train_eval_steps_per_second": 1.285,
"train_epoch": 1.0, "val_eval_loss": 0.6926347613334656, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 2.8758,
"val_eval_samples_per_second": 34.425, "val_eval_steps_per_second": 0.348,
"val_epoch": 1.0, "test_eval_loss": 0.6901114583015442, "test_eval_accuracy":
0.5434782608695652, "test_eval_precision": 0.5409836065573771,
"test_eval_recall": 1.0, "test_eval_f1": 0.7021276595744681,
"test_eval_runtime": 3.1784, "test_eval_samples_per_second": 57.891,
"test_eval_steps_per_second": 0.629, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:55:49.775844-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_pos_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,

```

```

"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rop_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rop_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 1.7063590288162231, "train_eval_accuracy":
0.5003295978905735, "train_eval_precision": 0.5, "train_eval_recall": 1.0,
"train_eval_f1": 0.6666666666666666, "train_eval_runtime": 4.8725,
"train_eval_samples_per_second": 311.342, "train_eval_steps_per_second": 2.463,
"train_epoch": 1.0, "val_eval_loss": 1.6273876428604126, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 1.6461,
"val_eval_samples_per_second": 60.142, "val_eval_steps_per_second": 0.607,
"val_epoch": 1.0, "test_eval_loss": 1.5908089876174927, "test_eval_accuracy":
0.5380434782608695, "test_eval_precision": 0.5380434782608695,
"test_eval_recall": 1.0, "test_eval_f1": 0.6996466431095406,
"test_eval_runtime": 1.7057, "test_eval_samples_per_second": 107.877,
"test_eval_steps_per_second": 1.173, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:56:45.349837-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-large", "learning_rate": 5e-06, "epochs":

```



```

1, "batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_pos_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.27.attn_norm.weight", "model.layers.27.attn.Wqkv.weight",
"model.layers.27.attn.Wo.weight", "model.layers.27.mlp_norm.weight",
"model.layers.27.mlp.Wi.weight", "model.layers.27.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 395833346, "model_trainable_params": 13309954,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-large",
"_commit_hash": "45bb4654a4d5aaff24dd11d4781fa46d39bf8c13",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 1024, "intermediate_size": 2624,
"num_hidden_layers": 28, "num_attention_heads": 16, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.6946010589599609, "train_eval_accuracy":
0.5385629531970996, "train_eval_precision": 0.5551330798479087,
"train_eval_recall": 0.38522427440633245, "train_eval_f1": 0.45482866043613707,

```

```

"train_eval_runtime": 9.9844, "train_eval_samples_per_second": 151.937,
"train_eval_steps_per_second": 1.202, "train_epoch": 1.0, "val_eval_loss":
0.7137817740440369, "val_eval_accuracy": 0.46464646464646464,
"val_eval_precision": 0.475, "val_eval_recall": 0.37254901960784315,
"val_eval_f1": 0.4175824175824176, "val_eval_runtime": 2.296,
"val_eval_samples_per_second": 43.118, "val_eval_steps_per_second": 0.436,
"val_epoch": 1.0, "test_eval_loss": 0.6987304091453552, "test_eval_accuracy":
0.4945652173913043, "test_eval_precision": 0.5454545454545454,
"test_eval_recall": 0.36363636363636365, "test_eval_f1": 0.43636363636363634,
"test_eval_runtime": 3.2754, "test_eval_samples_per_second": 56.177,
"test_eval_steps_per_second": 0.611, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:57:10.847139-07:00", "experiment_meta":
{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_pos_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",
"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",
"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,

```

```

"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":
"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":
null, "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":
"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":
12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,
"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.7006112337112427, "train_eval_accuracy":
0.5042847725774555, "train_eval_precision": 0.6, "train_eval_recall":
0.023746701846965697, "train_eval_f1": 0.04568527918781726,
"train_eval_runtime": 3.9408, "train_eval_samples_per_second": 384.946,
"train_eval_steps_per_second": 3.045, "train_epoch": 1.0, "val_eval_loss":
0.7038478851318359, "val_eval_accuracy": 0.48484848484848486,
"val_eval_precision": 0.5, "val_eval_recall": 0.0392156862745098, "val_eval_f1":
0.07272727272727272, "val_eval_runtime": 1.5929, "val_eval_samples_per_second":
62.15, "val_eval_steps_per_second": 0.628, "val_epoch": 1.0, "test_eval_loss":
0.7067509293556213, "test_eval_accuracy": 0.47282608695652173,
"test_eval_precision": 0.75, "test_eval_recall": 0.030303030303030304,
"test_eval_f1": 0.05825242718446602, "test_eval_runtime": 1.6117,
"test_eval_samples_per_second": 114.166, "test_eval_steps_per_second": 1.241,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:59:07.833763-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":

```

```

1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7281028032302856, "train_eval_accuracy": 0.4996704021094265,
"train_eval_precision": 0.4996704021094265, "train_eval_recall": 1.0,
"train_eval_f1": 0.6663736263736264, "train_eval_runtime": 7.106,
"train_eval_samples_per_second": 213.481, "train_eval_steps_per_second": 1.689,
"train_epoch": 1.0, "val_eval_loss": 0.7170491814613342, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 5.565,
"val_eval_samples_per_second": 17.79, "val_eval_steps_per_second": 0.18,
"val_epoch": 1.0, "test_eval_loss": 0.7107948064804077, "test_eval_accuracy":
0.5380434782608695, "test_eval_precision": 0.5380434782608695,
"test_eval_recall": 1.0, "test_eval_f1": 0.6996466431095406,
"test_eval_runtime": 5.2179, "test_eval_samples_per_second": 35.263,
"test_eval_steps_per_second": 0.383, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T08:59:54.445522-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",

```

```

"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6932850480079651,
"train_eval_accuracy": 0.4996704021094265, "train_eval_precision":
0.4996704021094265, "train_eval_recall": 1.0, "train_eval_f1":
0.6663736263736264, "train_eval_runtime": 6.9628,
"train_eval_samples_per_second": 217.872, "train_eval_steps_per_second": 1.723,
"train_epoch": 1.0, "val_eval_loss": 0.6924635171890259, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 5.0764,
"val_eval_samples_per_second": 19.502, "val_eval_steps_per_second": 0.197,
"val_epoch": 1.0, "test_eval_loss": 0.6913613677024841, "test_eval_accuracy":

```

```

0.5380434782608695, "test_eval_precision": 0.5380434782608695,
"test_eval_recall": 1.0, "test_eval_f1": 0.6996466431095406,
"test_eval_runtime": 5.1436, "test_eval_samples_per_second": 35.773,
"test_eval_steps_per_second": 0.389, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:02:32.514805-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaef65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.7652206420898438,
"train_eval_accuracy": 0.5201054713249835, "train_eval_precision":
0.556390977443609, "train_eval_recall": 0.19525065963060687, "train_eval_f1":
0.2890625, "train_eval_runtime": 11.9734, "train_eval_samples_per_second":
126.698, "train_eval_steps_per_second": 1.002, "train_epoch": 1.0,
"val_eval_loss": 0.7925207614898682, "val_eval_accuracy": 0.494949494949495,

```

```

"val_eval_precision": 0.5333333333333333, "val_eval_recall": 0.1568627450980392,
"val_eval_f1": 0.24242424242424243, "val_eval_runtime": 5.6379,
"val_eval_samples_per_second": 17.56, "val_eval_steps_per_second": 0.177,
"val_epoch": 1.0, "test_eval_loss": 0.7933067679405212, "test_eval_accuracy":
0.4891304347826087, "test_eval_precision": 0.5862068965517241,
"test_eval_recall": 0.1717171717171717, "test_eval_f1": 0.265625,
"test_eval_runtime": 5.8924, "test_eval_samples_per_second": 31.227,
"test_eval_steps_per_second": 0.339, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:03:38.728078-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},

```

```

"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdp", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6929916143417358,
"train_eval_accuracy": 0.4996704021094265, "train_eval_precision":
0.4996704021094265, "train_eval_recall": 1.0, "train_eval_f1":
0.6663736263736264, "train_eval_runtime": 11.9866,
"train_eval_samples_per_second": 126.558, "train_eval_steps_per_second": 1.001,
"train_epoch": 1.0, "val_eval_loss": 0.6942740678787231, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 5.7975,
"val_eval_samples_per_second": 17.076, "val_eval_steps_per_second": 0.172,
"val_epoch": 1.0, "test_eval_loss": 0.6895093321800232, "test_eval_accuracy":
0.5380434782608695, "test_eval_precision": 0.5380434782608695,
"test_eval_recall": 1.0, "test_eval_f1": 0.6996466431095406,
"test_eval_runtime": 6.1215, "test_eval_samples_per_second": 30.058,
"test_eval_steps_per_second": 0.327, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:04:48.486754-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,

```



```

"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 2.092895984649658, "train_eval_accuracy":
0.4996704021094265, "train_eval_precision": 0.4996704021094265,
"train_eval_recall": 1.0, "train_eval_f1": 0.6663736263736264,
"train_eval_runtime": 9.4985, "train_eval_samples_per_second": 159.709,
"train_eval_steps_per_second": 1.263, "train_epoch": 1.0, "val_eval_loss":
1.9611859321594238, "val_eval_accuracy": 0.5151515151515151,
"val_eval_precision": 0.5151515151515151, "val_eval_recall": 1.0, "val_eval_f1":
0.68, "val_eval_runtime": 5.988, "val_eval_samples_per_second": 16.533,
"val_eval_steps_per_second": 0.167, "val_epoch": 1.0, "test_eval_loss":
1.9127166271209717, "test_eval_accuracy": 0.5380434782608695,
"test_eval_precision": 0.5380434782608695, "test_eval_recall": 1.0,
"test_eval_f1": 0.6996466431095406, "test_eval_runtime": 5.4207,
"test_eval_samples_per_second": 33.944, "test_eval_steps_per_second": 0.369,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:06:02.927579-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-large", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.27.attn_norm.weight", "model.layers.27.attn.Wqkv.weight",
"model.layers.27.attn.Wo.weight", "model.layers.27.mlp_norm.weight",
"model.layers.27.mlp.Wi.weight", "model.layers.27.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":

```

```

{"model_total_params": 395833346, "model_trainable_params": 13309954,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-large",
"_commit_hash": "45bb4654a4d5aaff24dd11d4781fa46d39bf8c13",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 1024, "intermediate_size": 2624,
"num_hidden_layers": 28, "num_attention_heads": 16, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.6908855438232422, "train_eval_accuracy":
0.5504284772577456, "train_eval_precision": 0.5616883116883117,
"train_eval_recall": 0.45646437994722955, "train_eval_f1": 0.5036390101892285,
"train_eval_runtime": 13.8193, "train_eval_samples_per_second": 109.774,
"train_eval_steps_per_second": 0.868, "train_epoch": 1.0, "val_eval_loss":
0.6933701634407043, "val_eval_accuracy": 0.5353535353535354,
"val_eval_precision": 0.5531914893617021, "val_eval_recall": 0.5098039215686274,
"val_eval_f1": 0.5306122448979592, "val_eval_runtime": 5.6848,
"val_eval_samples_per_second": 17.415, "val_eval_steps_per_second": 0.176,
"val_epoch": 1.0, "test_eval_loss": 0.6932283043861389, "test_eval_accuracy":

```

```

0.5054347826086957, "test_eval_precision": 0.55, "test_eval_recall":
0.4444444444444444, "test_eval_f1": 0.49162011173184356, "test_eval_runtime":
6.0405, "test_eval_samples_per_second": 30.461, "test_eval_steps_per_second":
0.331, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:07:17.467972-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_dep_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7489381432533264, "train_eval_accuracy": 0.5003295978905735,
"train_eval_precision": 0.0, "train_eval_recall": 0.0, "train_eval_f1": 0.0,
"train_eval_runtime": 3.663, "train_eval_samples_per_second": 414.14,
"train_eval_steps_per_second": 3.276, "train_epoch": 1.0, "val_eval_loss":
0.7599362730979919, "val_eval_accuracy": 0.48484848484848486,
"val_eval_precision": 0.0, "val_eval_recall": 0.0, "val_eval_f1": 0.0,
"val_eval_runtime": 1.4558, "val_eval_samples_per_second": 68.005,
"val_eval_steps_per_second": 0.687, "val_epoch": 1.0, "test_eval_loss":

```

```

0.7753816843032837, "test_eval_accuracy": 0.46195652173913043,
"test_eval_precision": 0.0, "test_eval_recall": 0.0, "test_eval_f1": 0.0,
"test_eval_runtime": 2.1074, "test_eval_samples_per_second": 87.311,
"test_eval_steps_per_second": 0.949, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:07:38.008146-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_dep_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":

```

```

"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6934992074966431,
"train_eval_accuracy": 0.4996704021094265, "train_eval_precision":
0.4996704021094265, "train_eval_recall": 1.0, "train_eval_f1":
0.6663736263736264, "train_eval_runtime": 3.5564,
"train_eval_samples_per_second": 426.554, "train_eval_steps_per_second": 3.374,
"train_epoch": 1.0, "val_eval_loss": 0.693295419216156, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 1.7316,
"val_eval_samples_per_second": 57.173, "val_eval_steps_per_second": 0.578,
"val_epoch": 1.0, "test_eval_loss": 0.6918212175369263, "test_eval_accuracy":
0.5380434782608695, "test_eval_precision": 0.5380434782608695,
"test_eval_recall": 1.0, "test_eval_f1": 0.6996466431095406,
"test_eval_runtime": 1.8468, "test_eval_samples_per_second": 99.632,
"test_eval_steps_per_second": 1.083, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:09:31.843709-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_dep_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",

```

```

"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaeff65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.7167028188705444,
"train_eval_accuracy": 0.5266974291364536, "train_eval_precision":
0.6639344262295082, "train_eval_recall": 0.10686015831134564, "train_eval_f1":
0.18409090909090908, "train_eval_runtime": 8.0117,
"train_eval_samples_per_second": 189.348, "train_eval_steps_per_second": 1.498,
"train_epoch": 1.0, "val_eval_loss": 0.7302045822143555, "val_eval_accuracy":
0.5050505050505051, "val_eval_precision": 0.6, "val_eval_recall":
0.11764705882352941, "val_eval_f1": 0.19672131147540983, "val_eval_runtime":
1.7281, "val_eval_samples_per_second": 57.289, "val_eval_steps_per_second":
0.579, "val_epoch": 1.0, "test_eval_loss": 0.7514451742172241,
"test_eval_accuracy": 0.4782608695652174, "test_eval_precision": 0.6,
"test_eval_recall": 0.09090909090909091, "test_eval_f1": 0.15789473684210525,
"test_eval_runtime": 2.1826, "test_eval_samples_per_second": 84.303,
"test_eval_steps_per_second": 0.916, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:10:13.326257-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_dep_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,

```

```

"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6931219696998596,
"train_eval_accuracy": 0.4996704021094265, "train_eval_precision":
0.4996704021094265, "train_eval_recall": 1.0, "train_eval_f1":
0.6663736263736264, "train_eval_runtime": 8.4645,
"train_eval_samples_per_second": 179.219, "train_eval_steps_per_second": 1.418,
"train_epoch": 1.0, "val_eval_loss": 0.6944130659103394, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 2.0705,
"val_eval_samples_per_second": 47.814, "val_eval_steps_per_second": 0.483,
"val_epoch": 1.0, "test_eval_loss": 0.6902158856391907, "test_eval_accuracy":
0.5380434782608695, "test_eval_precision": 0.5380434782608695,
"test_eval_recall": 1.0, "test_eval_f1": 0.6996466431095406,
"test_eval_runtime": 2.4095, "test_eval_samples_per_second": 76.365,
"test_eval_steps_per_second": 0.83, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:10:56.497551-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_dep_seq", "y_col": "binary_complexity", "layers_to_unfreeze":

```

```
[
  "model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
  "model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
  "model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
  "model.final_norm.weight", "head.dense.weight", "head.norm.weight",
  "classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 2.8125100135803223, "train_eval_accuracy":
0.4996704021094265, "train_eval_precision": 0.4996704021094265,
"train_eval_recall": 1.0, "train_eval_f1": 0.6663736263736264,
"train_eval_runtime": 4.0818, "train_eval_samples_per_second": 371.649,
"train_eval_steps_per_second": 2.94, "train_epoch": 1.0, "val_eval_loss":
```



```

2.816037893295288, "val_eval_accuracy": 0.5151515151515151,
"val_eval_precision": 0.5151515151515151, "val_eval_recall": 1.0, "val_eval_f1":
0.68, "val_eval_runtime": 1.194, "val_eval_samples_per_second": 82.911,
"val_eval_steps_per_second": 0.837, "val_epoch": 1.0, "test_eval_loss":
2.5637753009796143, "test_eval_accuracy": 0.5380434782608695,
"test_eval_precision": 0.5380434782608695, "test_eval_recall": 1.0,
"test_eval_f1": 0.6996466431095406, "test_eval_runtime": 1.649,
"test_eval_samples_per_second": 111.58, "test_eval_steps_per_second": 1.213,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:11:50.113159-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-large", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_dep_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.27.attn_norm.weight", "model.layers.27.attn.Wqkv.weight",
"model.layers.27.attn.Wo.weight", "model.layers.27.mlp_norm.weight",
"model.layers.27.mlp.Wi.weight", "model.layers.27.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 395833346, "model_trainable_params": 13309954,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-large",
"_commit_hash": "45bb4654a4d5aaff24dd11d4781fa46d39bf8c13",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 1024, "intermediate_size": 2624,
"num_hidden_layers": 28, "num_attention_heads": 16, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,

```

```

"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.821459949016571, "train_eval_accuracy":
0.4996704021094265, "train_eval_precision": 0.4996699669966997,
"train_eval_recall": 0.9986807387862797, "train_eval_f1": 0.666080070391553,
"train_eval_runtime": 9.7858, "train_eval_samples_per_second": 155.021,
"train_eval_steps_per_second": 1.226, "train_epoch": 1.0, "val_eval_loss":
0.7944267988204956, "val_eval_accuracy": 0.5151515151515151,
"val_eval_precision": 0.5151515151515151, "val_eval_recall": 1.0, "val_eval_f1":
0.68, "val_eval_runtime": 2.1182, "val_eval_samples_per_second": 46.737,
"val_eval_steps_per_second": 0.472, "val_epoch": 1.0, "test_eval_loss":
0.8093674778938293, "test_eval_accuracy": 0.5380434782608695,
"test_eval_precision": 0.5380434782608695, "test_eval_recall": 1.0,
"test_eval_f1": 0.6996466431095406, "test_eval_runtime": 2.7709,
"test_eval_samples_per_second": 66.404, "test_eval_steps_per_second": 0.722,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:12:24.413752-07:00", "experiment_meta":
{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_dep_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",
"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",
"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,

```

```

"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":
"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":
null, "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":
"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":
12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,
"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.6945787668228149, "train_eval_accuracy":
0.4990112063282795, "train_eval_precision": 0.4954954954954955,
"train_eval_recall": 0.14511873350923482, "train_eval_f1": 0.22448979591836735,
"train_eval_runtime": 4.8336, "train_eval_samples_per_second": 313.848,
"train_eval_steps_per_second": 2.483, "train_epoch": 1.0, "val_eval_loss":
0.6965318918228149, "val_eval_accuracy": 0.45454545454545453,
"val_eval_precision": 0.38461538461538464, "val_eval_recall":
0.09803921568627451, "val_eval_f1": 0.15625, "val_eval_runtime": 2.0058,
"val_eval_samples_per_second": 49.356, "val_eval_steps_per_second": 0.499,
"val_epoch": 1.0, "test_eval_loss": 0.6993821859359741, "test_eval_accuracy":
0.483695652173913, "test_eval_precision": 0.5833333333333334,
"test_eval_recall": 0.1414141414141414, "test_eval_f1": 0.22764227642276422,
"test_eval_runtime": 2.7624, "test_eval_samples_per_second": 66.609,
"test_eval_steps_per_second": 0.724, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:13:49.588386-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_dep_seq", "y_col": "binary_complexity", "layers_to_unfreeze":

```

```
[
  "transformer.layer.11.rel_attn.q", "transformer.layer.11.rel_attn.k",
  "transformer.layer.11.rel_attn.v", "transformer.layer.11.rel_attn.o",
  "transformer.layer.11.rel_attn.r", "transformer.layer.11.rel_attn.r_r_bias",
  "transformer.layer.11.rel_attn.r_s_bias",
  "transformer.layer.11.rel_attn.r_w_bias",
  "transformer.layer.11.rel_attn.seg_embed",
  "transformer.layer.11.rel_attn.layer_norm.weight",
  "transformer.layer.11.rel_attn.layer_norm.bias",
  "transformer.layer.11.ff.layer_norm.weight",
  "transformer.layer.11.ff.layer_norm.bias",
  "transformer.layer.11.ff.layer_1.weight",
  "transformer.layer.11.ff.layer_1.bias",
  "transformer.layer.11.ff.layer_2.weight",
  "transformer.layer.11.ff.layer_2.bias", "sequence_summary.summary.weight",
  "sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
{"model_details": {"model_total_params": 117310466, "model_trainable_params":
8270594, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 768, "n_layer": 12, "n_head": 12, "d_head": 64,
"ff_activation": "gelu", "d_inner": 3072, "untie_r": true, "attn_type": "bi",
"initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1, "mem_len":
null, "reuse_len": null, "bi_data": false, "clamp_len": -1, "same_length":
false, "summary_type": "last", "summary_use_proj": true, "summary_activation":
"tanh", "summary_last_dropout": 0.1, "start_n_top": 5, "end_n_top": 5,
"bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2, "use_mems_eval": true,
"use_mems_train": false, "return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-base-cased",
"_commit_hash": "ceaa69c7bc5e512b5007106a7ccbb7daf24b2c79",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}, "run_metrics":
```

```

{"train_eval_loss": 0.704359769821167, "train_eval_accuracy":
0.4779169413315755, "train_eval_precision": 0.4734375, "train_eval_recall":
0.3997361477572559, "train_eval_f1": 0.4334763948497854, "train_eval_runtime":
5.2617, "train_eval_samples_per_second": 288.312, "train_eval_steps_per_second":
2.281, "train_epoch": 1.0, "val_eval_loss": 0.7084683775901794,
"val_eval_accuracy": 0.5151515151515151, "val_eval_precision":
0.5454545454545454, "val_eval_recall": 0.35294117647058826, "val_eval_f1":
0.42857142857142855, "val_eval_runtime": 1.7109, "val_eval_samples_per_second":
57.865, "val_eval_steps_per_second": 0.584, "val_epoch": 1.0, "test_eval_loss":
0.6974037885665894, "test_eval_accuracy": 0.5271739130434783,
"test_eval_precision": 0.5857142857142857, "test_eval_recall":
0.41414141414141414, "test_eval_f1": 0.48520710059171596, "test_eval_runtime":
2.2054, "test_eval_samples_per_second": 83.433, "test_eval_steps_per_second":
0.907, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:14:39.531630-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_dep_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.23.rel_attn.q", "transformer.layer.23.rel_attn.k",
"transformer.layer.23.rel_attn.v", "transformer.layer.23.rel_attn.o",
"transformer.layer.23.rel_attn.r", "transformer.layer.23.rel_attn.r_r_bias",
"transformer.layer.23.rel_attn.r_s_bias",
"transformer.layer.23.rel_attn.r_w_bias",
"transformer.layer.23.rel_attn.seg_embed",
"transformer.layer.23.rel_attn.layer_norm.weight",
"transformer.layer.23.rel_attn.layer_norm.bias",
"transformer.layer.23.ff.layer_norm.weight",
"transformer.layer.23.ff.layer_norm.bias",
"transformer.layer.23.ff.layer_1.weight",
"transformer.layer.23.ff.layer_1.bias",
"transformer.layer.23.ff.layer_2.weight",
"transformer.layer.23.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 361320450, "model_trainable_params":
14697474, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 1024, "n_layer": 24, "n_head": 16, "d_head":
64, "ff_activation": "gelu", "d_inner": 4096, "untie_r": true, "attn_type":
"bi", "initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1,
"mem_len": null, "reuse_len": null, "bi_data": false, "clamp_len": -1,
"same_length": false, "summary_type": "last", "summary_use_proj": true,
"summary_activation": "tanh", "summary_last_dropout": 0.1, "start_n_top": 5,
"end_n_top": 5, "bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2,
"use_mems_eval": true, "use_mems_train": false, "return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":

```

```

20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-large-cased",
"_commit_hash": "37658b4f179eaf971d127d16bfb9ca676a93034",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}, "run_metrics":
{"train_eval_loss": 0.6716889142990112, "train_eval_accuracy":
0.6025049439683586, "train_eval_precision": 0.6333907056798623,
"train_eval_recall": 0.48548812664907653, "train_eval_f1": 0.549663928304705,
"train_eval_runtime": 12.1895, "train_eval_samples_per_second": 124.451,
"train_eval_steps_per_second": 0.984, "train_epoch": 1.0, "val_eval_loss":
0.6995196342468262, "val_eval_accuracy": 0.5656565656565656,
"val_eval_precision": 0.6176470588235294, "val_eval_recall": 0.4117647058823529,
"val_eval_f1": 0.49411764705882355, "val_eval_runtime": 2.7012,
"val_eval_samples_per_second": 36.651, "val_eval_steps_per_second": 0.37,
"val_epoch": 1.0, "test_eval_loss": 0.6983924508094788, "test_eval_accuracy":
0.5217391304347826, "test_eval_precision": 0.5873015873015873,
"test_eval_recall": 0.37373737373737376, "test_eval_f1": 0.4567901234567901,
"test_eval_runtime": 3.4321, "test_eval_samples_per_second": 53.611,
"test_eval_steps_per_second": 0.583, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:17:08.972111-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_complexity_value", "y_col": "binary_complexity",
"layers_to_unfreeze": ["bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 7680002, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":

```

```

1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6918671727180481, "train_eval_accuracy": 0.5095583388266315,
"train_eval_precision": 0.5047554347826086, "train_eval_recall":
0.9802110817941952, "train_eval_f1": 0.6663677130044843, "train_eval_runtime":
7.0462, "train_eval_samples_per_second": 215.294, "train_eval_steps_per_second":
1.703, "train_epoch": 1.0, "val_eval_loss": 0.6908776164054871,
"val_eval_accuracy": 0.5151515151515151, "val_eval_precision":
0.5157894736842106, "val_eval_recall": 0.9607843137254902, "val_eval_f1":
0.6712328767123288, "val_eval_runtime": 6.0951, "val_eval_samples_per_second":
16.243, "val_eval_steps_per_second": 0.164, "val_epoch": 1.0, "test_eval_loss":
0.6944844722747803, "test_eval_accuracy": 0.5108695652173914,
"test_eval_precision": 0.5254237288135594, "test_eval_recall":
0.9393939393939394, "test_eval_f1": 0.6739130434782609, "test_eval_runtime":
5.1999, "test_eval_samples_per_second": 35.385, "test_eval_steps_per_second":
0.385, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:17:56.202178-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_complexity_value", "y_col": "binary_complexity",
"layers_to_unfreeze": ["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",

```

```

"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6937929391860962,
"train_eval_accuracy": 0.4996704021094265, "train_eval_precision":
0.4996704021094265, "train_eval_recall": 1.0, "train_eval_f1":
0.6663736263736264, "train_eval_runtime": 6.9958,
"train_eval_samples_per_second": 216.843, "train_eval_steps_per_second": 1.715,
"train_epoch": 1.0, "val_eval_loss": 0.6930475831031799, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 5.0636,
"val_eval_samples_per_second": 19.551, "val_eval_steps_per_second": 0.197,

```



```

"val_epoch": 1.0, "test_eval_loss": 0.6915158629417419, "test_eval_accuracy":
0.5380434782608695, "test_eval_precision": 0.5380434782608695,
"test_eval_recall": 1.0, "test_eval_f1": 0.6996466431095406,
"test_eval_runtime": 5.0485, "test_eval_samples_per_second": 36.446,
"test_eval_steps_per_second": 0.396, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:19:56.121628-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_complexity_value", "y_col": "binary_complexity",
"layers_to_unfreeze": ["bert.encoder.layer.23.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 333581314,
"model_trainable_params": 13647874, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaeff65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.7835530638694763,
"train_eval_accuracy": 0.5003295978905735, "train_eval_precision": 0.0,
"train_eval_recall": 0.0, "train_eval_f1": 0.0, "train_eval_runtime": 11.8937,
"train_eval_samples_per_second": 127.546, "train_eval_steps_per_second": 1.009,

```

```

"train_epoch": 1.0, "val_eval_loss": 0.7986159920692444, "val_eval_accuracy":
0.48484848484848486, "val_eval_precision": 0.0, "val_eval_recall": 0.0,
"val_eval_f1": 0.0, "val_eval_runtime": 5.3704, "val_eval_samples_per_second":
18.434, "val_eval_steps_per_second": 0.186, "val_epoch": 1.0, "test_eval_loss":
0.8214293718338013, "test_eval_accuracy": 0.46195652173913043,
"test_eval_precision": 0.0, "test_eval_recall": 0.0, "test_eval_f1": 0.0,
"test_eval_runtime": 5.7162, "test_eval_samples_per_second": 32.189,
"test_eval_steps_per_second": 0.35, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:21:00.819195-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_complexity_value", "y_col": "binary_complexity",
"layers_to_unfreeze": ["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},

```

```

"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdp", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6943228840827942,
"train_eval_accuracy": 0.4996704021094265, "train_eval_precision":
0.4996704021094265, "train_eval_recall": 1.0, "train_eval_f1":
0.6663736263736264, "train_eval_runtime": 12.0124,
"train_eval_samples_per_second": 126.287, "train_eval_steps_per_second": 0.999,
"train_epoch": 1.0, "val_eval_loss": 0.6942838430404663, "val_eval_accuracy":
0.5151515151515151, "val_eval_precision": 0.5151515151515151, "val_eval_recall":
1.0, "val_eval_f1": 0.68, "val_eval_runtime": 5.394,
"val_eval_samples_per_second": 18.354, "val_eval_steps_per_second": 0.185,
"val_epoch": 1.0, "test_eval_loss": 0.6916723847389221, "test_eval_accuracy":
0.5380434782608695, "test_eval_precision": 0.5380434782608695,
"test_eval_recall": 1.0, "test_eval_f1": 0.6996466431095406,
"test_eval_runtime": 5.8205, "test_eval_samples_per_second": 31.612,
"test_eval_steps_per_second": 0.344, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:22:08.126507-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_complexity_value", "y_col": "binary_complexity",
"layers_to_unfreeze": ["model.layers.21.attn_norm.weight",
"model.layers.21.attn.Wqkv.weight", "model.layers.21.attn.Wo.weight",
"model.layers.21.mlp_norm.weight", "model.layers.21.mlp.Wi.weight",
"model.layers.21.mlp.Wo.weight", "model.final_norm.weight", "head.dense.weight",
"head.norm.weight", "classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,

```

```

"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 2.284294843673706, "train_eval_accuracy":
0.5023071852340145, "train_eval_precision": 0.500992720052945,
"train_eval_recall": 0.9986807387862797, "train_eval_f1": 0.6672542970471573,
"train_eval_runtime": 7.9201, "train_eval_samples_per_second": 191.539,
"train_eval_steps_per_second": 1.515, "train_epoch": 1.0, "val_eval_loss":
2.441112756729126, "val_eval_accuracy": 0.5151515151515151,
"val_eval_precision": 0.5151515151515151, "val_eval_recall": 1.0, "val_eval_f1":
0.68, "val_eval_runtime": 5.1719, "val_eval_samples_per_second": 19.142,
"val_eval_steps_per_second": 0.193, "val_epoch": 1.0, "test_eval_loss":
2.0650463104248047, "test_eval_accuracy": 0.5380434782608695,
"test_eval_precision": 0.5380434782608695, "test_eval_recall": 1.0,
"test_eval_f1": 0.6996466431095406, "test_eval_runtime": 5.381,
"test_eval_samples_per_second": 34.194, "test_eval_steps_per_second": 0.372,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:23:18.687475-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-large", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_complexity_value", "y_col": "binary_complexity",
"layers_to_unfreeze": ["model.layers.27.attn_norm.weight",
"model.layers.27.attn.Wqkv.weight", "model.layers.27.attn.Wo.weight",
"model.layers.27.mlp_norm.weight", "model.layers.27.mlp.Wi.weight",
"model.layers.27.mlp.Wo.weight", "model.final_norm.weight", "head.dense.weight",
"head.norm.weight", "classifier.weight", "classifier.bias"]}, "model_details":

```

```

{"model_total_params": 395833346, "model_trainable_params": 13309954,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-large",
"_commit_hash": "45bb4654a4d5aaff24dd11d4781fa46d39bf8c13",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 1024, "intermediate_size": 2624,
"num_hidden_layers": 28, "num_attention_heads": 16, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.7531933784484863, "train_eval_accuracy":
0.5293342122610415, "train_eval_precision": 0.5158273381294964,
"train_eval_recall": 0.945910290237467, "train_eval_f1": 0.6675977653631285,
"train_eval_runtime": 13.0496, "train_eval_samples_per_second": 116.249,
"train_eval_steps_per_second": 0.92, "train_epoch": 1.0, "val_eval_loss":
0.7499406337738037, "val_eval_accuracy": 0.5454545454545454,
"val_eval_precision": 0.5319148936170213, "val_eval_recall": 0.9803921568627451,
"val_eval_f1": 0.6896551724137931, "val_eval_runtime": 5.3536,
"val_eval_samples_per_second": 18.492, "val_eval_steps_per_second": 0.187,
"val_epoch": 1.0, "test_eval_loss": 0.7823100090026855, "test_eval_accuracy":

```

```

0.5217391304347826, "test_eval_precision": 0.5321637426900585,
"test_eval_recall": 0.9191919191919192, "test_eval_f1": 0.674074074074074,
"test_eval_runtime": 5.9642, "test_eval_samples_per_second": 30.851,
"test_eval_steps_per_second": 0.335, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:24:09.162968-07:00", "experiment_meta":
{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_complexity_value", "y_col": "binary_complexity",
"layers_to_unfreeze":
["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",
"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",
"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,

```

```

"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":
"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":
null, "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":
"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":
12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,
"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.6947001218795776, "train_eval_accuracy":
0.5082399472643375, "train_eval_precision": 0.5229007633587787,
"train_eval_recall": 0.18073878627968337, "train_eval_f1": 0.26862745098039215,
"train_eval_runtime": 8.0303, "train_eval_samples_per_second": 188.91,
"train_eval_steps_per_second": 1.494, "train_epoch": 1.0, "val_eval_loss":
0.6897398829460144, "val_eval_accuracy": 0.5454545454545454,
"val_eval_precision": 0.7142857142857143, "val_eval_recall":
0.19607843137254902, "val_eval_f1": 0.3076923076923077, "val_eval_runtime":
5.3719, "val_eval_samples_per_second": 18.429, "val_eval_steps_per_second":
0.186, "val_epoch": 1.0, "test_eval_loss": 0.7092766165733337,
"test_eval_accuracy": 0.41847826086956524, "test_eval_precision":
0.36666666666666664, "test_eval_recall": 0.11111111111111111, "test_eval_f1":
0.17054263565891473, "test_eval_runtime": 5.2695,
"test_eval_samples_per_second": 34.918, "test_eval_steps_per_second": 0.38,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:25:58.961947-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_complexity_value", "y_col": "binary_complexity",
"layers_to_unfreeze": ["transformer.layer.11.rel_attn.q",
"transformer.layer.11.rel_attn.k", "transformer.layer.11.rel_attn.v",
"transformer.layer.11.rel_attn.o", "transformer.layer.11.rel_attn.r",
"transformer.layer.11.rel_attn.r_r_bias",
"transformer.layer.11.rel_attn.r_s_bias",
"transformer.layer.11.rel_attn.r_w_bias",
"transformer.layer.11.rel_attn.seg_embed",
"transformer.layer.11.rel_attn.layer_norm.weight",
"transformer.layer.11.rel_attn.layer_norm.bias",
"transformer.layer.11.ff.layer_norm.weight",
"transformer.layer.11.ff.layer_norm.bias",
"transformer.layer.11.ff.layer_1.weight",
"transformer.layer.11.ff.layer_1.bias",
"transformer.layer.11.ff.layer_2.weight",
"transformer.layer.11.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]}},

```

```

"model_details": {"model_total_params": 117310466, "model_trainable_params":
8270594, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 768, "n_layer": 12, "n_head": 12, "d_head": 64,
"ff_activation": "gelu", "d_inner": 3072, "untie_r": true, "attn_type": "bi",
"initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1, "mem_len":
null, "reuse_len": null, "bi_data": false, "clamp_len": -1, "same_length":
false, "summary_type": "last", "summary_use_proj": true, "summary_activation":
"tanh", "summary_last_dropout": 0.1, "start_n_top": 5, "end_n_top": 5,
"bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2, "use_mems_eval": true,
"use_mems_train": false, "return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-base-cased",
"_commit_hash": "ceaa69c7bc5e512b5007106a7ccbb7daf24b2c79",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}}, "run_metrics":
{"train_eval_loss": 0.7431444525718689, "train_eval_accuracy":
0.4976928147659855, "train_eval_precision": 0.4986737400530504,
"train_eval_recall": 0.9920844327176781, "train_eval_f1": 0.6637246248896734,
"train_eval_runtime": 8.5212, "train_eval_samples_per_second": 178.027,
"train_eval_steps_per_second": 1.408, "train_epoch": 1.0, "val_eval_loss":
0.7216587066650391, "val_eval_accuracy": 0.5151515151515151,
"val_eval_precision": 0.5151515151515151, "val_eval_recall": 1.0, "val_eval_f1":
0.68, "val_eval_runtime": 5.3723, "val_eval_samples_per_second": 18.428,
"val_eval_steps_per_second": 0.186, "val_epoch": 1.0, "test_eval_loss":
0.7248739004135132, "test_eval_accuracy": 0.5380434782608695,
"test_eval_precision": 0.5380434782608695, "test_eval_recall": 1.0,
"test_eval_f1": 0.6996466431095406, "test_eval_runtime": 5.7066,
"test_eval_samples_per_second": 32.243, "test_eval_steps_per_second": 0.35,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T09:27:14.071695-07:00", "experiment_meta":

```



```

{"model_name": "xlnet/xlnet-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_complexity_value", "y_col": "binary_complexity",
"layers_to_unfreeze": ["transformer.layer.23.rel_attn.q",
"transformer.layer.23.rel_attn.k", "transformer.layer.23.rel_attn.v",
"transformer.layer.23.rel_attn.o", "transformer.layer.23.rel_attn.r",
"transformer.layer.23.rel_attn.r_r_bias",
"transformer.layer.23.rel_attn.r_s_bias",
"transformer.layer.23.rel_attn.r_w_bias",
"transformer.layer.23.rel_attn.seg_embed",
"transformer.layer.23.rel_attn.layer_norm.weight",
"transformer.layer.23.rel_attn.layer_norm.bias",
"transformer.layer.23.ff.layer_norm.weight",
"transformer.layer.23.ff.layer_norm.bias",
"transformer.layer.23.ff.layer_1.weight",
"transformer.layer.23.ff.layer_1.bias",
"transformer.layer.23.ff.layer_2.weight",
"transformer.layer.23.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 361320450, "model_trainable_params":
14697474, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 1024, "n_layer": 24, "n_head": 16, "d_head":
64, "ff_activation": "gelu", "d_inner": 4096, "untie_r": true, "attn_type":
"bi", "initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1,
"mem_len": null, "reuse_len": null, "bi_data": false, "clamp_len": -1,
"same_length": false, "summary_type": "last", "summary_use_proj": true,
"summary_activation": "tanh", "summary_last_dropout": 0.1, "start_n_top": 5,
"end_n_top": 5, "bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2,
"use_mems_eval": true, "use_mems_train": false, "return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":

```

```

"single_label_classification", "_name_or_path": "xlnet/xlnet-large-cased",
"_commit_hash": "37658b4f179eaf971d127d16bfb9ca676a93034",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}}, "run_metrics":
{"train_eval_loss": 0.6809659600257874, "train_eval_accuracy":
0.5695451549110085, "train_eval_precision": 0.5670498084291188,
"train_eval_recall": 0.5857519788918206, "train_eval_f1": 0.5762491888384166,
"train_eval_runtime": 15.2501, "train_eval_samples_per_second": 99.475,
"train_eval_steps_per_second": 0.787, "train_epoch": 1.0, "val_eval_loss":
0.6860107779502869, "val_eval_accuracy": 0.5555555555555556,
"val_eval_precision": 0.5813953488372093, "val_eval_recall":
0.49019607843137253, "val_eval_f1": 0.5319148936170213, "val_eval_runtime":
5.7372, "val_eval_samples_per_second": 17.256, "val_eval_steps_per_second":
0.174, "val_epoch": 1.0, "test_eval_loss": 0.6885326504707336,
"test_eval_accuracy": 0.5815217391304348, "test_eval_precision":
0.6170212765957447, "test_eval_recall": 0.5858585858585859, "test_eval_f1":
0.6010362694300518, "test_eval_runtime": 6.1535, "test_eval_samples_per_second":
29.902, "test_eval_steps_per_second": 0.325, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:13:09.859416-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",

```

```

"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7107220888137817, "train_eval_accuracy": 0.505, "train_eval_precision":
0.6666666666666666, "train_eval_recall": 0.0005770340450086555, "train_eval_f1":
0.0011530700490054772, "train_eval_runtime": 11.6121,
"train_eval_samples_per_second": 602.821, "train_eval_steps_per_second": 4.736,
"train_epoch": 1.0, "val_eval_loss": 0.7034121155738831, "val_eval_accuracy":
0.519, "val_eval_precision": 1.0, "val_eval_recall": 0.002074688796680498,
"val_eval_f1": 0.004140786749482402, "val_eval_runtime": 2.5231,
"val_eval_samples_per_second": 396.341, "val_eval_steps_per_second": 3.171,
"val_epoch": 1.0, "test_eval_loss": 0.7061905860900879, "test_eval_accuracy":
0.519, "test_eval_precision": 1.0, "test_eval_recall": 0.002074688796680498,
"test_eval_f1": 0.004140786749482402, "test_eval_runtime": 2.5783,
"test_eval_samples_per_second": 387.854, "test_eval_steps_per_second": 3.103,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:13:53.720769-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,

```

```

"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6930952668190002,
"train_eval_accuracy": 0.5012857142857143, "train_eval_precision":
0.48847926267281105, "train_eval_recall": 0.1529140219272937, "train_eval_f1":
0.23291584267194024, "train_eval_runtime": 11.6927,
"train_eval_samples_per_second": 598.664, "train_eval_steps_per_second": 4.704,
"train_epoch": 1.0, "val_eval_loss": 0.692793607711792, "val_eval_accuracy":
0.523, "val_eval_precision": 0.5161290322580645, "val_eval_recall":
0.16597510373443983, "val_eval_f1": 0.25117739403453687, "val_eval_runtime":
2.8054, "val_eval_samples_per_second": 356.451, "val_eval_steps_per_second":
2.852, "val_epoch": 1.0, "test_eval_loss": 0.6929181814193726,
"test_eval_accuracy": 0.526, "test_eval_precision": 0.5263157894736842,
"test_eval_recall": 0.16597510373443983, "test_eval_f1": 0.25236593059936907,
"test_eval_runtime": 2.5638, "test_eval_samples_per_second": 390.049,
"test_eval_steps_per_second": 3.12, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:16:36.591148-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":

```

```

null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaeff65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6925804018974304,
"train_eval_accuracy": 0.5231428571428571, "train_eval_precision":
0.5191387559808612, "train_eval_recall": 0.500865551067513, "train_eval_f1":
0.5098384728340676, "train_eval_runtime": 32.7239,
"train_eval_samples_per_second": 213.911, "train_eval_steps_per_second": 1.681,
"train_epoch": 1.0, "val_eval_loss": 0.6903700232505798, "val_eval_accuracy":
0.529, "val_eval_precision": 0.5116279069767442, "val_eval_recall":
0.5020746887966805, "val_eval_f1": 0.506806282722513, "val_eval_runtime":
6.0213, "val_eval_samples_per_second": 166.077, "val_eval_steps_per_second":
1.329, "val_epoch": 1.0, "test_eval_loss": 0.6916303634643555,
"test_eval_accuracy": 0.504, "test_eval_precision": 0.48478260869565215,
"test_eval_recall": 0.46265560165975106, "test_eval_f1": 0.4734607218683652,
"test_eval_runtime": 5.6817, "test_eval_samples_per_second": 176.003,
"test_eval_steps_per_second": 1.408, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:18:27.432408-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",

```

```

"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6934049725532532,
"train_eval_accuracy": 0.4948571428571429, "train_eval_precision":
0.4948194197750148, "train_eval_recall": 0.9645124062319677, "train_eval_f1":

```

```

0.6540794365094893, "train_eval_runtime": 32.8241,
"train_eval_samples_per_second": 213.258, "train_eval_steps_per_second": 1.676,
"train_epoch": 1.0, "val_eval_loss": 0.6938955187797546, "val_eval_accuracy":
0.485, "val_eval_precision": 0.48297213622291024, "val_eval_recall":
0.970954356846473, "val_eval_f1": 0.6450723638869745, "val_eval_runtime":
5.7301, "val_eval_samples_per_second": 174.518, "val_eval_steps_per_second":
1.396, "val_epoch": 1.0, "test_eval_loss": 0.6937005519866943,
"test_eval_accuracy": 0.481, "test_eval_precision": 0.48074921956295524,
"test_eval_recall": 0.9585062240663901, "test_eval_f1": 0.6403326403326404,
"test_eval_runtime": 5.6423, "test_eval_samples_per_second": 177.231,
"test_eval_steps_per_second": 1.418, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:19:43.354538-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,

```

```

"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 1.2918223142623901, "train_eval_accuracy":
0.4972857142857143, "train_eval_precision": 0.4958844541077807,
"train_eval_recall": 0.9212348528563186, "train_eval_f1": 0.6447248864209995,
"train_eval_runtime": 15.116, "train_eval_samples_per_second": 463.087,
"train_eval_steps_per_second": 3.639, "train_epoch": 1.0, "val_eval_loss":
1.301515817642212, "val_eval_accuracy": 0.483, "val_eval_precision":
0.4807480748074808, "val_eval_recall": 0.9066390041493776, "val_eval_f1":
0.6283249460819554, "val_eval_runtime": 3.5945, "val_eval_samples_per_second":
278.199, "val_eval_steps_per_second": 2.226, "val_epoch": 1.0, "test_eval_loss":
1.3263109922409058, "test_eval_accuracy": 0.492, "test_eval_precision":
0.48580786026200873, "test_eval_recall": 0.9232365145228216, "test_eval_f1":
0.6366237482117311, "test_eval_runtime": 3.0909, "test_eval_samples_per_second":
323.529, "test_eval_steps_per_second": 2.588, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:22:01.137549-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-large", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.27.attn_norm.weight", "model.layers.27.attn.Wqkv.weight",
"model.layers.27.attn.Wo.weight", "model.layers.27.mlp_norm.weight",
"model.layers.27.mlp.Wi.weight", "model.layers.27.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 395833346, "model_trainable_params": 13309954,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":

```



```
[{"ModernBertForSequenceClassification"}, {"finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-large",
"_commit_hash": "45bb4654a4d5aaff24dd11d4781fa46d39bf8c13",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 1024, "intermediate_size": 2624,
"num_hidden_layers": 28, "num_attention_heads": 16, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, {"run_metrics":
{"train_eval_loss": 0.7279731035232544, "train_eval_accuracy": 0.501,
"train_eval_precision": 0.49565776777098747, "train_eval_recall":
0.4446047316791691, "train_eval_f1": 0.46874524714828897, "train_eval_runtime":
39.1327, "train_eval_samples_per_second": 178.879,
"train_eval_steps_per_second": 1.405, "train_epoch": 1.0, "val_eval_loss":
0.7094017267227173, "val_eval_accuracy": 0.534, "val_eval_precision":
0.5192307692307693, "val_eval_recall": 0.44813278008298757, "val_eval_f1":
0.48106904231625836, "val_eval_runtime": 6.8595, "val_eval_samples_per_second":
145.783, "val_eval_steps_per_second": 1.166, "val_epoch": 1.0, "test_eval_loss":
0.7202410697937012, "test_eval_accuracy": 0.496, "test_eval_precision":
0.47453703703703703, "test_eval_recall": 0.42531120331950206, "test_eval_f1":
0.4485776805251641, "test_eval_runtime": 6.5347, "test_eval_samples_per_second":
153.029, "test_eval_steps_per_second": 1.224, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:23:09.924365-07:00", "experiment_meta":
{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
```

```

"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",
"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",
"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":
"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":
null, "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":
"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":
12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,
"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.6951717138290405, "train_eval_accuracy":
0.49442857142857144, "train_eval_precision": 0.4870152970473141,
"train_eval_recall": 0.3949798038084247, "train_eval_f1": 0.43619563485741597,
"train_eval_runtime": 14.2405, "train_eval_samples_per_second": 491.555,
"train_eval_steps_per_second": 3.862, "train_epoch": 1.0, "val_eval_loss":

```

```

0.6954110264778137, "val_eval_accuracy": 0.495, "val_eval_precision":
0.472289156626506, "val_eval_recall": 0.4066390041493776, "val_eval_f1":
0.43701226309921964, "val_eval_runtime": 3.0071, "val_eval_samples_per_second":
332.548, "val_eval_steps_per_second": 2.66, "val_epoch": 1.0, "test_eval_loss":
0.6917496919631958, "test_eval_accuracy": 0.521, "test_eval_precision":
0.503562945368171, "test_eval_recall": 0.43983402489626555, "test_eval_f1":
0.4695459579180509, "test_eval_runtime": 2.9578, "test_eval_samples_per_second":
338.087, "test_eval_steps_per_second": 2.705, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:25:06.081078-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.11.rel_attn.q", "transformer.layer.11.rel_attn.k",
"transformer.layer.11.rel_attn.v", "transformer.layer.11.rel_attn.o",
"transformer.layer.11.rel_attn.r", "transformer.layer.11.rel_attn.r_r_bias",
"transformer.layer.11.rel_attn.r_s_bias",
"transformer.layer.11.rel_attn.r_w_bias",
"transformer.layer.11.rel_attn.seg_embed",
"transformer.layer.11.rel_attn.layer_norm.weight",
"transformer.layer.11.rel_attn.layer_norm.bias",
"transformer.layer.11.ff.layer_norm.weight",
"transformer.layer.11.ff.layer_norm.bias",
"transformer.layer.11.ff.layer_1.weight",
"transformer.layer.11.ff.layer_1.bias",
"transformer.layer.11.ff.layer_2.weight",
"transformer.layer.11.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 117310466, "model_trainable_params":
8270594, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 768, "n_layer": 12, "n_head": 12, "d_head": 64,
"ff_activation": "gelu", "d_inner": 3072, "untie_r": true, "attn_type": "bi",
"initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1, "mem_len":
null, "reuse_len": null, "bi_data": false, "clamp_len": -1, "same_length":
false, "summary_type": "last", "summary_use_proj": true, "summary_activation":
"tanh", "summary_last_dropout": 0.1, "start_n_top": 5, "end_n_top": 5,
"bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2, "use_mems_eval": true,
"use_mems_train": false, "return_dict": true, "output_hidden_states": false,
"output attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,

```

```

"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-base-cased",
"_commit_hash": "ceaa69c7bc5e512b5007106a7ccbb7daf24b2c79",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}}, "run_metrics":
{"train_eval_loss": 0.6971307992935181, "train_eval_accuracy":
0.5152857142857142, "train_eval_precision": 0.5076761303890641,
"train_eval_recall": 0.6964800923254472, "train_eval_f1": 0.5872764870453716,
"train_eval_runtime": 17.4367, "train_eval_samples_per_second": 401.452,
"train_eval_steps_per_second": 3.154, "train_epoch": 1.0, "val_eval_loss":
0.6957322955131531, "val_eval_accuracy": 0.51, "val_eval_precision":
0.49404761904761907, "val_eval_recall": 0.6887966804979253, "val_eval_f1":
0.5753899480069324, "val_eval_runtime": 3.3935, "val_eval_samples_per_second":
294.682, "val_eval_steps_per_second": 2.357, "val_epoch": 1.0, "test_eval_loss":
0.6914323568344116, "test_eval_accuracy": 0.53, "test_eval_precision":
0.5089285714285714, "test_eval_recall": 0.7095435684647303, "test_eval_f1":
0.5927209705372617, "test_eval_runtime": 3.5345, "test_eval_samples_per_second":
282.926, "test_eval_steps_per_second": 2.263, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:26:53.073858-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},

```

```

"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdp", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6929013729095459, "train_eval_accuracy": 0.5167142857142857,
"train_eval_precision": 0.5152854511970534, "train_eval_recall":
0.40363531448355455, "train_eval_f1": 0.45267756026532924, "train_eval_runtime":
12.0498, "train_eval_samples_per_second": 580.922,
"train_eval_steps_per_second": 4.564, "train_epoch": 1.0, "val_eval_loss":
0.6926511526107788, "val_eval_accuracy": 0.499, "val_eval_precision":
0.4743935309973046, "val_eval_recall": 0.3651452282157676, "val_eval_f1":
0.4126611957796014, "val_eval_runtime": 4.2108, "val_eval_samples_per_second":
237.482, "val_eval_steps_per_second": 1.9, "val_epoch": 1.0, "test_eval_loss":
0.6949717998504639, "test_eval_accuracy": 0.505, "test_eval_precision":
0.4844124700239808, "test_eval_recall": 0.4190871369294606, "test_eval_f1":
0.44938820912124583, "test_eval_runtime": 4.5378,
"test_eval_samples_per_second": 220.373, "test_eval_steps_per_second": 1.763,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:27:38.894103-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",

```

```

"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6929454803466797,
"train_eval_accuracy": 0.5095714285714286, "train_eval_precision":
0.5690376569037657, "train_eval_recall": 0.03923831506058857, "train_eval_f1":
0.07341430499325236, "train_eval_runtime": 11.7477,
"train_eval_samples_per_second": 595.862, "train_eval_steps_per_second": 4.682,
"train_epoch": 1.0, "val_eval_loss": 0.6928278207778931, "val_eval_accuracy":
0.517, "val_eval_precision": 0.48717948717948717, "val_eval_recall":
0.03941908713692946, "val_eval_f1": 0.07293666026871401, "val_eval_runtime":
2.9604, "val_eval_samples_per_second": 337.79, "val_eval_steps_per_second":
2.702, "val_epoch": 1.0, "test_eval_loss": 0.6927818655967712,
"test_eval_accuracy": 0.524, "test_eval_precision": 0.5714285714285714,
"test_eval_recall": 0.04979253112033195, "test_eval_f1": 0.0916030534351145,
"test_eval_runtime": 3.2921, "test_eval_samples_per_second": 303.759,
"test_eval_steps_per_second": 2.43, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:31:18.460141-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":

```

```

"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."], "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaeff65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6948090195655823,
"train_eval_accuracy": 0.49457142857142855, "train_eval_precision":
0.488134475939354, "train_eval_recall": 0.4272937103289094, "train_eval_f1":
0.4556923076923077, "train_eval_runtime": 33.6986,
"train_eval_samples_per_second": 207.724, "train_eval_steps_per_second": 1.632,
"train_epoch": 1.0, "val_eval_loss": 0.6934207677841187, "val_eval_accuracy":
0.504, "val_eval_precision": 0.4830097087378641, "val_eval_recall":
0.41286307053941906, "val_eval_f1": 0.4451901565995526, "val_eval_runtime":
6.4497, "val_eval_samples_per_second": 155.046, "val_eval_steps_per_second":
1.24, "val_epoch": 1.0, "test_eval_loss": 0.6967467069625854,
"test_eval_accuracy": 0.471, "test_eval_precision": 0.44719101123595506,
"test_eval_recall": 0.41286307053941906, "test_eval_f1": 0.42934196332254587,
"test_eval_runtime": 6.4582, "test_eval_samples_per_second": 154.842,

```

```

"test_eval_steps_per_second": 1.239, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:33:17.189772-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,

```



```

"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6932507157325745,
"train_eval_accuracy": 0.5007142857142857, "train_eval_precision":
0.4977487967706878, "train_eval_recall": 0.9249855741488748, "train_eval_f1":
0.6472191379832442, "train_eval_runtime": 33.114,
"train_eval_samples_per_second": 211.391, "train_eval_steps_per_second": 1.661,
"train_epoch": 1.0, "val_eval_loss": 0.6939289569854736, "val_eval_accuracy":
0.485, "val_eval_precision": 0.48184818481848185, "val_eval_recall":
0.9087136929460581, "val_eval_f1": 0.6297627606038821, "val_eval_runtime":
5.861, "val_eval_samples_per_second": 170.62, "val_eval_steps_per_second":
1.365, "val_epoch": 1.0, "test_eval_loss": 0.6932868361473083,
"test_eval_accuracy": 0.492, "test_eval_precision": 0.4857456140350877,
"test_eval_recall": 0.9190871369294605, "test_eval_f1": 0.6355810616929699,
"test_eval_runtime": 5.909, "test_eval_samples_per_second": 169.234,
"test_eval_steps_per_second": 1.354, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:34:40.079843-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":

```

```

null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.9788786172866821, "train_eval_accuracy":
0.4908571428571429, "train_eval_precision": 0.49174250084260196,
"train_eval_recall": 0.8418926716676284, "train_eval_f1": 0.6208510638297873,
"train_eval_runtime": 15.7614, "train_eval_samples_per_second": 444.124,
"train_eval_steps_per_second": 3.49, "train_epoch": 1.0, "val_eval_loss":
0.9877772331237793, "val_eval_accuracy": 0.501, "val_eval_precision":
0.4899645808736718, "val_eval_recall": 0.8609958506224067, "val_eval_f1":
0.6245297215951844, "val_eval_runtime": 3.2005, "val_eval_samples_per_second":
312.452, "val_eval_steps_per_second": 2.5, "val_epoch": 1.0, "test_eval_loss":
0.9952330589294434, "test_eval_accuracy": 0.482, "test_eval_precision":
0.4784688995215311, "test_eval_recall": 0.8298755186721992, "test_eval_f1":
0.6069802731411229, "test_eval_runtime": 3.1505, "test_eval_samples_per_second":
317.405, "test_eval_steps_per_second": 2.539, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:37:07.881839-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-large", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.27.attn_norm.weight", "model.layers.27.attn.Wqkv.weight",
"model.layers.27.attn.Wo.weight", "model.layers.27.mlp_norm.weight",
"model.layers.27.mlp.Wi.weight", "model.layers.27.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 395833346, "model_trainable_params": 13309954,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":

```

```

1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-large",
"_commit_hash": "45bb4654a4d5aaff24dd11d4781fa46d39bf8c13",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 1024, "intermediate_size": 2624,
"num_hidden_layers": 28, "num_attention_heads": 16, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.6959407329559326, "train_eval_accuracy":
0.5201428571428571, "train_eval_precision": 0.5189649060616802,
"train_eval_recall": 0.42238892094633584, "train_eval_f1": 0.46572292031175444,
"train_eval_runtime": 39.3291, "train_eval_samples_per_second": 177.985,
"train_eval_steps_per_second": 1.398, "train_epoch": 1.0, "val_eval_loss":
0.6924781203269958, "val_eval_accuracy": 0.548, "val_eval_precision":
0.5360576923076923, "val_eval_recall": 0.46265560165975106, "val_eval_f1":
0.49665924276169265, "val_eval_runtime": 6.5215, "val_eval_samples_per_second":
153.338, "val_eval_steps_per_second": 1.227, "val_epoch": 1.0, "test_eval_loss":
0.6966729760169983, "test_eval_accuracy": 0.537, "test_eval_precision":
0.5232273838630807, "test_eval_recall": 0.44398340248962653, "test_eval_f1":
0.48035914702581367, "test_eval_runtime": 6.6455,
"test_eval_samples_per_second": 150.478, "test_eval_steps_per_second": 1.204,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:38:20.139725-07:00", "experiment_meta":
{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"morph_sequence", "y_col": "binary_complexity", "layers_to_unfreeze":

```

```

["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",
"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",
"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":
"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":
null, "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":
"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":
12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,

```

```

"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.6930907368659973, "train_eval_accuracy":
0.5065714285714286, "train_eval_precision": 0.5063694267515924,
"train_eval_recall": 0.13762261973456433, "train_eval_f1": 0.21642468239564427,
"train_eval_runtime": 14.699, "train_eval_samples_per_second": 476.222,
"train_eval_steps_per_second": 3.742, "train_epoch": 1.0, "val_eval_loss":
0.6929465532302856, "val_eval_accuracy": 0.51, "val_eval_precision":
0.47435897435897434, "val_eval_recall": 0.15352697095435686, "val_eval_f1":
0.23197492163009403, "val_eval_runtime": 3.4754, "val_eval_samples_per_second":
287.739, "val_eval_steps_per_second": 2.302, "val_epoch": 1.0, "test_eval_loss":
0.6920635104179382, "test_eval_accuracy": 0.522, "test_eval_precision":
0.5133333333333333, "test_eval_recall": 0.15975103734439833, "test_eval_f1":
0.24367088607594936, "test_eval_runtime": 3.7738,
"test_eval_samples_per_second": 264.982, "test_eval_steps_per_second": 2.12,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:39:56.481694-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_pos_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",

```

```

"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6869339942932129, "train_eval_accuracy": 0.5546153846153846,
"train_eval_precision": 0.5769230769230769, "train_eval_recall":
0.33070866141732286, "train_eval_f1": 0.42042042042042044, "train_eval_runtime":
3.2392, "train_eval_samples_per_second": 401.332, "train_eval_steps_per_second":
3.396, "train_epoch": 1.0, "val_eval_loss": 0.6866680979728699,
"val_eval_accuracy": 0.564, "val_eval_precision": 0.49333333333333335,
"val_eval_recall": 0.3425925925925926, "val_eval_f1": 0.40437158469945356,
"val_eval_runtime": 1.8481, "val_eval_samples_per_second": 135.276,
"val_eval_steps_per_second": 1.082, "val_epoch": 1.0, "test_eval_loss":
0.6892813444137573, "test_eval_accuracy": 0.54, "test_eval_precision": 0.59375,
"test_eval_recall": 0.2992125984251969, "test_eval_f1": 0.39790575916230364,
"test_eval_runtime": 2.4734, "test_eval_samples_per_second": 101.074,
"test_eval_steps_per_second": 0.809, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:40:16.950685-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_pos_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,

```

```

"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6941998600959778,
"train_eval_accuracy": 0.48846153846153845, "train_eval_precision":
0.48846153846153845, "train_eval_recall": 1.0, "train_eval_f1":
0.6563307493540051, "train_eval_runtime": 3.1734,
"train_eval_samples_per_second": 409.652, "train_eval_steps_per_second": 3.466,
"train_epoch": 1.0, "val_eval_loss": 0.6972572803497314, "val_eval_accuracy":
0.432, "val_eval_precision": 0.432, "val_eval_recall": 1.0, "val_eval_f1":
0.6033519553072626, "val_eval_runtime": 1.6836, "val_eval_samples_per_second":
148.495, "val_eval_steps_per_second": 1.188, "val_epoch": 1.0, "test_eval_loss":
0.6936155557632446, "test_eval_accuracy": 0.508, "test_eval_precision": 0.508,
"test_eval_recall": 1.0, "test_eval_f1": 0.6737400530503979,
"test_eval_runtime": 2.1264, "test_eval_samples_per_second": 117.57,
"test_eval_steps_per_second": 0.941, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:41:00.589088-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_complexity_value", "y_col": "binary_complexity",
"layers_to_unfreeze": ["bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 7680002, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,

```

```

"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7295464873313904, "train_eval_accuracy": 0.42923076923076925,
"train_eval_precision": 0.4521038495971352, "train_eval_recall":
0.7952755905511811, "train_eval_f1": 0.5764840182648402, "train_eval_runtime":
3.2475, "train_eval_samples_per_second": 400.305, "train_eval_steps_per_second":
3.387, "train_epoch": 1.0, "val_eval_loss": 0.7467932105064392,
"val_eval_accuracy": 0.392, "val_eval_precision": 0.39814814814814814,
"val_eval_recall": 0.7962962962962963, "val_eval_f1": 0.5308641975308642,
"val_eval_runtime": 1.8081, "val_eval_samples_per_second": 138.264,
"val_eval_steps_per_second": 1.106, "val_epoch": 1.0, "test_eval_loss":
0.7119725346565247, "test_eval_accuracy": 0.492, "test_eval_precision": 0.5,
"test_eval_recall": 0.8818897637795275, "test_eval_f1": 0.6381766381766382,
"test_eval_runtime": 1.6033, "test_eval_samples_per_second": 155.931,
"test_eval_steps_per_second": 1.247, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:41:21.557512-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_complexity_value", "y_col": "binary_complexity",
"layers_to_unfreeze": ["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",

```



```

"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6943135857582092,
"train_eval_accuracy": 0.48846153846153845, "train_eval_precision":
0.48846153846153845, "train_eval_recall": 1.0, "train_eval_f1":
0.6563307493540051, "train_eval_runtime": 3.0831,
"train_eval_samples_per_second": 421.656, "train_eval_steps_per_second": 3.568,
"train_epoch": 1.0, "val_eval_loss": 0.6973581314086914, "val_eval_accuracy":

```

```

0.432, "val_eval_precision": 0.432, "val_eval_recall": 1.0, "val_eval_f1":
0.6033519553072626, "val_eval_runtime": 1.499, "val_eval_samples_per_second":
166.775, "val_eval_steps_per_second": 1.334, "val_epoch": 1.0, "test_eval_loss":
0.6932892203330994, "test_eval_accuracy": 0.508, "test_eval_precision": 0.508,
"test_eval_recall": 1.0, "test_eval_f1": 0.6737400530503979,
"test_eval_runtime": 1.9134, "test_eval_samples_per_second": 130.657,
"test_eval_steps_per_second": 1.045, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:43:24.607114-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7385442852973938, "train_eval_accuracy": 0.5115384615384615,
"train_eval_precision": 0.0, "train_eval_recall": 0.0, "train_eval_f1": 0.0,
"train_eval_runtime": 6.7011, "train_eval_samples_per_second": 193.998,
"train_eval_steps_per_second": 1.642, "train_epoch": 1.0, "val_eval_loss":
0.7039474844932556, "val_eval_accuracy": 0.568, "val_eval_precision": 0.0,

```

```

"val_eval_recall": 0.0, "val_eval_f1": 0.0, "val_eval_runtime": 6.0312,
"val_eval_samples_per_second": 41.451, "val_eval_steps_per_second": 0.332,
"val_epoch": 1.0, "test_eval_loss": 0.7523830533027649, "test_eval_accuracy":
0.492, "test_eval_precision": 0.0, "test_eval_recall": 0.0, "test_eval_f1": 0.0,
"test_eval_runtime": 5.2613, "test_eval_samples_per_second": 47.516,
"test_eval_steps_per_second": 0.38, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:44:13.600081-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,

```

```

"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6938289999961853,
"train_eval_accuracy": 0.48846153846153845, "train_eval_precision":
0.48846153846153845, "train_eval_recall": 1.0, "train_eval_f1":
0.6563307493540051, "train_eval_runtime": 6.6613,
"train_eval_samples_per_second": 195.158, "train_eval_steps_per_second": 1.651,
"train_epoch": 1.0, "val_eval_loss": 0.6957210898399353, "val_eval_accuracy":
0.432, "val_eval_precision": 0.432, "val_eval_recall": 1.0, "val_eval_f1":
0.6033519553072626, "val_eval_runtime": 5.7688, "val_eval_samples_per_second":
43.337, "val_eval_steps_per_second": 0.347, "val_epoch": 1.0, "test_eval_loss":
0.6929977536201477, "test_eval_accuracy": 0.508, "test_eval_precision": 0.508,
"test_eval_recall": 1.0, "test_eval_f1": 0.6737400530503979,
"test_eval_runtime": 6.6631, "test_eval_samples_per_second": 37.52,
"test_eval_steps_per_second": 0.3, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:45:53.187635-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,

```

```

"task_specific_params": null, "problem_type": "single_label_classification",
"name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaeff65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.7558677792549133,
"train_eval_accuracy": 0.48846153846153845, "train_eval_precision":
0.48846153846153845, "train_eval_recall": 1.0, "train_eval_f1":
0.6563307493540051, "train_eval_runtime": 11.0849,
"train_eval_samples_per_second": 117.277, "train_eval_steps_per_second": 0.992,
"train_epoch": 1.0, "val_eval_loss": 0.7884328365325928, "val_eval_accuracy":
0.432, "val_eval_precision": 0.432, "val_eval_recall": 1.0, "val_eval_f1":
0.6033519553072626, "val_eval_runtime": 6.0527, "val_eval_samples_per_second":
41.304, "val_eval_steps_per_second": 0.33, "val_epoch": 1.0, "test_eval_loss":
0.7418806552886963, "test_eval_accuracy": 0.508, "test_eval_precision": 0.508,
"test_eval_recall": 1.0, "test_eval_f1": 0.6737400530503979,
"test_eval_runtime": 6.2108, "test_eval_samples_per_second": 40.252,
"test_eval_steps_per_second": 0.322, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:46:58.140135-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,

```

```

"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.694771409034729,
"train_eval_accuracy": 0.48846153846153845, "train_eval_precision":
0.48846153846153845, "train_eval_recall": 1.0, "train_eval_f1":
0.6563307493540051, "train_eval_runtime": 11.2276,
"train_eval_samples_per_second": 115.786, "train_eval_steps_per_second": 0.98,
"train_epoch": 1.0, "val_eval_loss": 0.7003488540649414, "val_eval_accuracy":
0.432, "val_eval_precision": 0.432, "val_eval_recall": 1.0, "val_eval_f1":
0.6033519553072626, "val_eval_runtime": 6.0652, "val_eval_samples_per_second":
41.219, "val_eval_steps_per_second": 0.33, "val_epoch": 1.0, "test_eval_loss":
0.6941280961036682, "test_eval_accuracy": 0.508, "test_eval_precision": 0.508,
"test_eval_recall": 1.0, "test_eval_f1": 0.6737400530503979,
"test_eval_runtime": 6.0572, "test_eval_samples_per_second": 41.273,
"test_eval_steps_per_second": 0.33, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:48:04.963948-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",

```

```

"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 1.793071985244751, "train_eval_accuracy":
0.48923076923076925, "train_eval_precision": 0.4888375673595073,
"train_eval_recall": 1.0, "train_eval_f1": 0.656670113753878,
"train_eval_runtime": 7.4268, "train_eval_samples_per_second": 175.043,
"train_eval_steps_per_second": 1.481, "train_epoch": 1.0, "val_eval_loss":
1.9776371717453003, "val_eval_accuracy": 0.432, "val_eval_precision": 0.432,

```

```

"val_eval_recall": 1.0, "val_eval_f1": 0.6033519553072626, "val_eval_runtime":
5.7829, "val_eval_samples_per_second": 43.231, "val_eval_steps_per_second":
0.346, "val_epoch": 1.0, "test_eval_loss": 1.7405925989151,
"test_eval_accuracy": 0.504, "test_eval_precision": 0.5060240963855421,
"test_eval_recall": 0.9921259842519685, "test_eval_f1": 0.6702127659574468,
"test_eval_runtime": 5.5729, "test_eval_samples_per_second": 44.86,
"test_eval_steps_per_second": 0.359, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:50:15.529764-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6922239065170288, "train_eval_accuracy": 0.52, "train_eval_precision":
0.5369127516778524, "train_eval_recall": 0.12598425196850394, "train_eval_f1":
0.20408163265306123, "train_eval_runtime": 6.7707,
"train_eval_samples_per_second": 192.004, "train_eval_steps_per_second": 1.625,
"train_epoch": 1.0, "val_eval_loss": 0.690558910369873, "val_eval_accuracy":

```



```

0.54, "val_eval_precision": 0.3870967741935484, "val_eval_recall":
0.1111111111111111, "val_eval_f1": 0.17266187050359713, "val_eval_runtime":
5.3881, "val_eval_samples_per_second": 46.399, "val_eval_steps_per_second":
0.371, "val_epoch": 1.0, "test_eval_loss": 0.6926782131195068,
"test_eval_accuracy": 0.536, "test_eval_precision": 0.6571428571428571,
"test_eval_recall": 0.18110236220472442, "test_eval_f1": 0.2839506172839506,
"test_eval_runtime": 5.694, "test_eval_samples_per_second": 43.906,
"test_eval_steps_per_second": 0.351, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:51:04.986964-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},

```

```

"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.694027304649353,
"train_eval_accuracy": 0.48846153846153845, "train_eval_precision":
0.48846153846153845, "train_eval_recall": 1.0, "train_eval_f1":
0.6563307493540051, "train_eval_runtime": 6.6433,
"train_eval_samples_per_second": 195.685, "train_eval_steps_per_second": 1.656,
"train_epoch": 1.0, "val_eval_loss": 0.6969093084335327, "val_eval_accuracy":
0.432, "val_eval_precision": 0.432, "val_eval_recall": 1.0, "val_eval_f1":
0.6033519553072626, "val_eval_runtime": 6.0109, "val_eval_samples_per_second":
41.591, "val_eval_steps_per_second": 0.333, "val_epoch": 1.0, "test_eval_loss":
0.6934626698493958, "test_eval_accuracy": 0.508, "test_eval_precision": 0.508,
"test_eval_recall": 1.0, "test_eval_f1": 0.6737400530503979,
"test_eval_runtime": 5.3, "test_eval_samples_per_second": 47.17,
"test_eval_steps_per_second": 0.377, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:52:59.586165-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},

```

```

"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaeff65", "_attn_implementation_internal":
"sdp", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.7090158462524414,
"train_eval_accuracy": 0.5161538461538462, "train_eval_precision":
0.6071428571428571, "train_eval_recall": 0.026771653543307086, "train_eval_f1":
0.05128205128205128, "train_eval_runtime": 10.7766,
"train_eval_samples_per_second": 120.631, "train_eval_steps_per_second": 1.021,
"train_epoch": 1.0, "val_eval_loss": 0.6827822327613831, "val_eval_accuracy":
0.556, "val_eval_precision": 0.0, "val_eval_recall": 0.0, "val_eval_f1": 0.0,
"val_eval_runtime": 6.4947, "val_eval_samples_per_second": 38.493,
"val_eval_steps_per_second": 0.308, "val_epoch": 1.0, "test_eval_loss":
0.7255847454071045, "test_eval_accuracy": 0.5, "test_eval_precision": 0.75,
"test_eval_recall": 0.023622047244094488, "test_eval_f1": 0.04580152671755725,
"test_eval_runtime": 6.1189, "test_eval_samples_per_second": 40.857,
"test_eval_steps_per_second": 0.327, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:54:07.902996-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",

```

```

"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6939055919647217,
"train_eval_accuracy": 0.48846153846153845, "train_eval_precision":
0.48846153846153845, "train_eval_recall": 1.0, "train_eval_f1":
0.6563307493540051, "train_eval_runtime": 11.1213,
"train_eval_samples_per_second": 116.893, "train_eval_steps_per_second": 0.989,
"train_epoch": 1.0, "val_eval_loss": 0.7001383304595947, "val_eval_accuracy":
0.432, "val_eval_precision": 0.432, "val_eval_recall": 1.0, "val_eval_f1":
0.6033519553072626, "val_eval_runtime": 6.3596, "val_eval_samples_per_second":
39.31, "val_eval_steps_per_second": 0.314, "val_epoch": 1.0, "test_eval_loss":
0.6923198103904724, "test_eval_accuracy": 0.508, "test_eval_precision": 0.508,
"test_eval_recall": 1.0, "test_eval_f1": 0.6737400530503979,
"test_eval_runtime": 6.3408, "test_eval_samples_per_second": 39.427,
"test_eval_steps_per_second": 0.315, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:55:17.508176-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":

```

```

"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 3.0217947959899902, "train_eval_accuracy":
0.48846153846153845, "train_eval_precision": 0.48846153846153845,
"train_eval_recall": 1.0, "train_eval_f1": 0.6563307493540051,
"train_eval_runtime": 7.8515, "train_eval_samples_per_second": 165.573,

```

```

"train_eval_steps_per_second": 1.401, "train_epoch": 1.0, "val_eval_loss":
3.3979573249816895, "val_eval_accuracy": 0.432, "val_eval_precision": 0.432,
"val_eval_recall": 1.0, "val_eval_f1": 0.6033519553072626, "val_eval_runtime":
6.4001, "val_eval_samples_per_second": 39.062, "val_eval_steps_per_second":
0.312, "val_epoch": 1.0, "test_eval_loss": 2.9320790767669678,
"test_eval_accuracy": 0.508, "test_eval_precision": 0.508, "test_eval_recall":
1.0, "test_eval_f1": 0.6737400530503979, "test_eval_runtime": 6.3905,
"test_eval_samples_per_second": 39.121, "test_eval_steps_per_second": 0.313,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:56:30.166015-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-large", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.27.attn_norm.weight", "model.layers.27.attn.Wqkv.weight",
"model.layers.27.attn.Wo.weight", "model.layers.27.mlp_norm.weight",
"model.layers.27.mlp.Wi.weight", "model.layers.27.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 395833346, "model_trainable_params": 13309954,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-large",
"_commit_hash": "45bb4654a4d5aaff24dd11d4781fa46d39bf8c13",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 1024, "intermediate_size": 2624,
"num_hidden_layers": 28, "num_attention_heads": 16, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,

```

```

"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.7140898704528809, "train_eval_accuracy":
0.48846153846153845, "train_eval_precision": 0.4858490566037736,
"train_eval_recall": 0.8110236220472441, "train_eval_f1": 0.6076696165191741,
"train_eval_runtime": 12.5934, "train_eval_samples_per_second": 103.228,
"train_eval_steps_per_second": 0.873, "train_epoch": 1.0, "val_eval_loss":
0.7463028430938721, "val_eval_accuracy": 0.444, "val_eval_precision":
0.418848167539267, "val_eval_recall": 0.7407407407407407, "val_eval_f1":
0.5351170568561873, "val_eval_runtime": 6.5125, "val_eval_samples_per_second":
38.388, "val_eval_steps_per_second": 0.307, "val_epoch": 1.0, "test_eval_loss":
0.6978663802146912, "test_eval_accuracy": 0.54, "test_eval_precision":
0.529126213592233, "test_eval_recall": 0.8582677165354331, "test_eval_f1":
0.6546546546546547, "test_eval_runtime": 6.6667, "test_eval_samples_per_second":
37.5, "test_eval_steps_per_second": 0.3, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T13:57:22.713721-07:00", "experiment_meta":
{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",
"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",
"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,

```

```

"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":
"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":
null, "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":
"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":
12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,
"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.6939628720283508, "train_eval_accuracy":
0.5046153846153846, "train_eval_precision": 0.40425531914893614,
"train_eval_recall": 0.029921259842519685, "train_eval_f1": 0.05571847507331378,
"train_eval_runtime": 7.9691, "train_eval_samples_per_second": 163.129,
"train_eval_steps_per_second": 1.38, "train_epoch": 1.0, "val_eval_loss":
0.687483012676239, "val_eval_accuracy": 0.552, "val_eval_precision": 0.375,
"val_eval_recall": 0.05555555555555555, "val_eval_f1": 0.0967741935483871,
"val_eval_runtime": 6.3335, "val_eval_samples_per_second": 39.473,
"val_eval_steps_per_second": 0.316, "val_epoch": 1.0, "test_eval_loss":
0.6991525888442993, "test_eval_accuracy": 0.476, "test_eval_precision": 0.4,
"test_eval_recall": 0.06299212598425197, "test_eval_f1": 0.10884353741496598,
"test_eval_runtime": 5.6771, "test_eval_samples_per_second": 44.037,
"test_eval_steps_per_second": 0.352, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:02:47.684987-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.11.rel_attn.q", "transformer.layer.11.rel_attn.k",
"transformer.layer.11.rel_attn.v", "transformer.layer.11.rel_attn.o",

```



```

"transformer.layer.11.rel_attn.r", "transformer.layer.11.rel_attn.r_r_bias",
"transformer.layer.11.rel_attn.r_s_bias",
"transformer.layer.11.rel_attn.r_w_bias",
"transformer.layer.11.rel_attn.seg_embed",
"transformer.layer.11.rel_attn.layer_norm.weight",
"transformer.layer.11.rel_attn.layer_norm.bias",
"transformer.layer.11.ff.layer_norm.weight",
"transformer.layer.11.ff.layer_norm.bias",
"transformer.layer.11.ff.layer_1.weight",
"transformer.layer.11.ff.layer_1.bias",
"transformer.layer.11.ff.layer_2.weight",
"transformer.layer.11.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 117310466, "model_trainable_params":
8270594, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 768, "n_layer": 12, "n_head": 12, "d_head": 64,
"ff_activation": "gelu", "d_inner": 3072, "untie_r": true, "attn_type": "bi",
"initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1, "mem_len":
null, "reuse_len": null, "bi_data": false, "clamp_len": -1, "same_length":
false, "summary_type": "last", "summary_use_proj": true, "summary_activation":
"tanh", "summary_last_dropout": 0.1, "start_n_top": 5, "end_n_top": 5,
"bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2, "use_mems_eval": true,
"use_mems_train": false, "return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-base-cased",
"_commit_hash": "ceaa69c7bc5e512b5007106a7ccbb7daf24b2c79",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}, "run_metrics":
{"train_eval_loss": 0.7101758718490601, "train_eval_accuracy":
0.5069230769230769, "train_eval_precision": 0.47058823529411764,

```

```

"train_eval_recall": 0.07559055118110236, "train_eval_f1": 0.13025780189959293,
"train_eval_runtime": 8.146, "train_eval_samples_per_second": 159.587,
"train_eval_steps_per_second": 1.35, "train_epoch": 1.0, "val_eval_loss":
0.6996338963508606, "val_eval_accuracy": 0.52, "val_eval_precision":
0.22727272727272727, "val_eval_recall": 0.046296296296296294, "val_eval_f1":
0.07692307692307693, "val_eval_runtime": 5.7982, "val_eval_samples_per_second":
43.117, "val_eval_steps_per_second": 0.345, "val_epoch": 1.0, "test_eval_loss":
0.7219932675361633, "test_eval_accuracy": 0.488, "test_eval_precision":
0.47058823529411764, "test_eval_recall": 0.06299212598425197, "test_eval_f1":
0.11111111111111111, "test_eval_runtime": 5.4212, "test_eval_samples_per_second":
46.115, "test_eval_steps_per_second": 0.369, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:04:00.698819-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_dep_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7031439542770386, "train_eval_accuracy": 0.48923076923076925,

```

```

"train_eval_precision": 0.4886983632112237, "train_eval_recall":
0.9874015748031496, "train_eval_f1": 0.6538060479666319, "train_eval_runtime":
7.0868, "train_eval_samples_per_second": 183.441, "train_eval_steps_per_second":
1.552, "train_epoch": 1.0, "val_eval_loss": 0.7093123197555542,
"val_eval_accuracy": 0.44, "val_eval_precision": 0.43548387096774194,
"val_eval_recall": 1.0, "val_eval_f1": 0.6067415730337079, "val_eval_runtime":
5.3126, "val_eval_samples_per_second": 47.058, "val_eval_steps_per_second":
0.376, "val_epoch": 1.0, "test_eval_loss": 0.6966758966445923,
"test_eval_accuracy": 0.504, "test_eval_precision": 0.5060240963855421,
"test_eval_recall": 0.9921259842519685, "test_eval_f1": 0.6702127659574468,
"test_eval_runtime": 5.1688, "test_eval_samples_per_second": 48.367,
"test_eval_steps_per_second": 0.387, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:04:47.233984-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_dep_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,

```

```

"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffb7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6939612627029419,
"train_eval_accuracy": 0.48846153846153845, "train_eval_precision":
0.48846153846153845, "train_eval_recall": 1.0, "train_eval_f1":
0.6563307493540051, "train_eval_runtime": 6.7508,
"train_eval_samples_per_second": 192.57, "train_eval_steps_per_second": 1.629,
"train_epoch": 1.0, "val_eval_loss": 0.6966370940208435, "val_eval_accuracy":
0.432, "val_eval_precision": 0.432, "val_eval_recall": 1.0, "val_eval_f1":
0.6033519553072626, "val_eval_runtime": 5.4377, "val_eval_samples_per_second":
45.975, "val_eval_steps_per_second": 0.368, "val_epoch": 1.0, "test_eval_loss":
0.6931424736976624, "test_eval_accuracy": 0.508, "test_eval_precision": 0.508,
"test_eval_recall": 1.0, "test_eval_f1": 0.6737400530503979,
"test_eval_runtime": 5.2957, "test_eval_samples_per_second": 47.208,
"test_eval_steps_per_second": 0.378, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:08:04.588653-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,

```

```

"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6924888491630554, "train_eval_accuracy": 0.5194285714285715,
"train_eval_precision": 0.5144968732234224, "train_eval_recall":
0.5222158107328332, "train_eval_f1": 0.518327605956472, "train_eval_runtime":
15.6598, "train_eval_samples_per_second": 447.005,
"train_eval_steps_per_second": 3.512, "train_epoch": 1.0, "val_eval_loss":
0.6895529627799988, "val_eval_accuracy": 0.529, "val_eval_precision":
0.510934393638171, "val_eval_recall": 0.533195020746888, "val_eval_f1":
0.5218274111675127, "val_eval_runtime": 6.1787, "val_eval_samples_per_second":
161.847, "val_eval_steps_per_second": 1.295, "val_epoch": 1.0, "test_eval_loss":
0.6907671093940735, "test_eval_accuracy": 0.527, "test_eval_precision":
0.5090180360721442, "test_eval_recall": 0.5269709543568465, "test_eval_f1":
0.5178389398572885, "test_eval_runtime": 6.2103, "test_eval_samples_per_second":
161.024, "test_eval_steps_per_second": 1.288, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:09:17.371645-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",

```

```

"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6928629875183105,
"train_eval_accuracy": 0.5088571428571429, "train_eval_precision":
0.5864197530864198, "train_eval_recall": 0.027409117137911138, "train_eval_f1":
0.05237045203969129, "train_eval_runtime": 16.0701,
"train_eval_samples_per_second": 435.592, "train_eval_steps_per_second": 3.423,
"train_epoch": 1.0, "val_eval_loss": 0.6925140619277954, "val_eval_accuracy":
0.525, "val_eval_precision": 0.6296296296296297, "val_eval_recall":
0.035269709543568464, "val_eval_f1": 0.06679764243614932, "val_eval_runtime":
6.1543, "val_eval_samples_per_second": 162.488, "val_eval_steps_per_second":
1.3, "val_epoch": 1.0, "test_eval_loss": 0.6926174163818359,
"test_eval_accuracy": 0.524, "test_eval_precision": 0.625, "test_eval_recall":
0.03112033195020747, "test_eval_f1": 0.05928853754940711, "test_eval_runtime":
6.2151, "test_eval_samples_per_second": 160.898, "test_eval_steps_per_second":
1.287, "test_epoch": 1.0}}

```

```

{"timestamp": "2025-04-11T14:12:27.890541-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaaff65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6985175609588623,
"train_eval_accuracy": 0.519, "train_eval_precision": 0.5258755880815473,
"train_eval_recall": 0.29024812463935373, "train_eval_f1": 0.37404722067298757,
"train_eval_runtime": 36.4598, "train_eval_samples_per_second": 191.992,
"train_eval_steps_per_second": 1.509, "train_epoch": 1.0, "val_eval_loss":
0.6884389519691467, "val_eval_accuracy": 0.548, "val_eval_precision":
0.5539568345323741, "val_eval_recall": 0.31950207468879666, "val_eval_f1":
0.4052631578947368, "val_eval_runtime": 10.2198, "val_eval_samples_per_second":
97.849, "val_eval_steps_per_second": 0.783, "val_epoch": 1.0, "test_eval_loss":
0.6968370079994202, "test_eval_accuracy": 0.529, "test_eval_precision":

```

```

0.5209125475285171, "test_eval_recall": 0.2842323651452282, "test_eval_f1":
0.36778523489932885, "test_eval_runtime": 9.2744,
"test_eval_samples_per_second": 107.824, "test_eval_steps_per_second": 0.863,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:14:47.113991-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":

```



```

"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6929659843444824,
"train_eval_accuracy": 0.5008571428571429, "train_eval_precision":
0.49781044729433843, "train_eval_recall": 0.9183496826312753, "train_eval_f1":
0.6456389452332657, "train_eval_runtime": 37.5884,
"train_eval_samples_per_second": 186.228, "train_eval_steps_per_second": 1.463,
"train_epoch": 1.0, "val_eval_loss": 0.6934168934822083, "val_eval_accuracy":
0.485, "val_eval_precision": 0.48164627363737483, "val_eval_recall":
0.8983402489626556, "val_eval_f1": 0.6270818247646633, "val_eval_runtime":
10.1521, "val_eval_samples_per_second": 98.502, "val_eval_steps_per_second":
0.788, "val_epoch": 1.0, "test_eval_loss": 0.6934801340103149,
"test_eval_accuracy": 0.485, "test_eval_precision": 0.4819277108433735,
"test_eval_recall": 0.9128630705394191, "test_eval_f1": 0.6308243727598566,
"test_eval_runtime": 10.0344, "test_eval_samples_per_second": 99.657,
"test_eval_steps_per_second": 0.797, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:16:29.436501-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":

```

```

{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 1.0902917385101318, "train_eval_accuracy":
0.4947142857142857, "train_eval_precision": 0.49487512631730907,
"train_eval_recall": 0.9890363531448355, "train_eval_f1": 0.6596747811026653,
"train_eval_runtime": 21.0544, "train_eval_samples_per_second": 332.472,
"train_eval_steps_per_second": 2.612, "train_epoch": 1.0, "val_eval_loss":
1.1110937595367432, "val_eval_accuracy": 0.482, "val_eval_precision":
0.4817813765182186, "val_eval_recall": 0.9875518672199171, "val_eval_f1":
0.6476190476190476, "val_eval_runtime": 7.4772, "val_eval_samples_per_second":
133.739, "val_eval_steps_per_second": 1.07, "val_epoch": 1.0, "test_eval_loss":
1.1363499164581299, "test_eval_accuracy": 0.482, "test_eval_precision":
0.48185483870967744, "test_eval_recall": 0.991701244813278, "test_eval_f1":
0.6485753052917232, "test_eval_runtime": 7.1919, "test_eval_samples_per_second":
139.045, "test_eval_steps_per_second": 1.112, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:17:35.472750-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_dep_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":

```

```

50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6964274048805237, "train_eval_accuracy": 0.5054285714285714,
"train_eval_precision": 0.5333333333333333, "train_eval_recall":
0.009232544720138488, "train_eval_f1": 0.018150879183210438,
"train_eval_runtime": 12.9498, "train_eval_samples_per_second": 540.551,
"train_eval_steps_per_second": 4.247, "train_epoch": 1.0, "val_eval_loss":
0.6955777406692505, "val_eval_accuracy": 0.514, "val_eval_precision": 0.3,
"val_eval_recall": 0.006224066390041493, "val_eval_f1": 0.012195121951219513,
"val_eval_runtime": 3.7279, "val_eval_samples_per_second": 268.25,
"val_eval_steps_per_second": 2.146, "val_epoch": 1.0, "test_eval_loss":
0.695040225982666, "test_eval_accuracy": 0.52, "test_eval_precision": 0.75,
"test_eval_recall": 0.006224066390041493, "test_eval_f1": 0.012345679012345678,
"test_eval_runtime": 2.9916, "test_eval_samples_per_second": 334.265,
"test_eval_steps_per_second": 2.674, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:18:21.788725-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_dep_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",

```

```

"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6930841207504272,
"train_eval_accuracy": 0.5052857142857143, "train_eval_precision":
0.5104895104895105, "train_eval_recall": 0.021061742642815927, "train_eval_f1":
0.04045441950678858, "train_eval_runtime": 11.5908,
"train_eval_samples_per_second": 603.929, "train_eval_steps_per_second": 4.745,
"train_epoch": 1.0, "val_eval_loss": 0.6927040815353394, "val_eval_accuracy":
0.517, "val_eval_precision": 0.4666666666666667, "val_eval_recall":
0.014522821576763486, "val_eval_f1": 0.028169014084507043, "val_eval_runtime":
3.0052, "val_eval_samples_per_second": 332.755, "val_eval_steps_per_second":

```

```

2.662, "val_epoch": 1.0, "test_eval_loss": 0.6929304003715515,
"test_eval_accuracy": 0.523, "test_eval_precision": 0.6086956521739131,
"test_eval_recall": 0.029045643153526972, "test_eval_f1": 0.055445544554455446,
"test_eval_runtime": 2.9254, "test_eval_samples_per_second": 341.829,
"test_eval_steps_per_second": 2.735, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:21:11.054443-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_dep_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaeff65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6922749876976013,
"train_eval_accuracy": 0.5217142857142857, "train_eval_precision":
0.525674499564839, "train_eval_recall": 0.3485285631852279, "train_eval_f1":
0.41915336571825124, "train_eval_runtime": 33.3486,
"train_eval_samples_per_second": 209.904, "train_eval_steps_per_second": 1.649,

```

```

"train_epoch": 1.0, "val_eval_loss": 0.6924840211868286, "val_eval_accuracy":
0.507, "val_eval_precision": 0.48328267477203646, "val_eval_recall":
0.32987551867219916, "val_eval_f1": 0.3921085080147966, "val_eval_runtime":
5.807, "val_eval_samples_per_second": 172.205, "val_eval_steps_per_second":
1.378, "val_epoch": 1.0, "test_eval_loss": 0.6916177868843079,
"test_eval_accuracy": 0.531, "test_eval_precision": 0.5203761755485894,
"test_eval_recall": 0.34439834024896265, "test_eval_f1": 0.41448189762796506,
"test_eval_runtime": 6.1339, "test_eval_samples_per_second": 163.027,
"test_eval_steps_per_second": 1.304, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:23:01.199916-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_dep_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":

```

```

"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6924881935119629,
"train_eval_accuracy": 0.524, "train_eval_precision": 0.5130910511918718,
"train_eval_recall": 0.7576457010963646, "train_eval_f1": 0.6118359739049394,
"train_eval_runtime": 33.0468, "train_eval_samples_per_second": 211.821,
"train_eval_steps_per_second": 1.664, "train_epoch": 1.0, "val_eval_loss":
0.6934066414833069, "val_eval_accuracy": 0.495, "val_eval_precision":
0.4847277556440903, "val_eval_recall": 0.7572614107883817, "val_eval_f1":
0.5910931174089069, "val_eval_runtime": 5.8582, "val_eval_samples_per_second":
170.702, "val_eval_steps_per_second": 1.366, "val_epoch": 1.0, "test_eval_loss":
0.6925405263900757, "test_eval_accuracy": 0.53, "test_eval_precision":
0.5084033613445378, "test_eval_recall": 0.7531120331950207, "test_eval_f1":
0.6070234113712375, "test_eval_runtime": 6.0578, "test_eval_samples_per_second":
165.077, "test_eval_steps_per_second": 1.321, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:24:18.521788-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_dep_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,

```

```

"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rop_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rop_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 1.6851955652236938, "train_eval_accuracy":
0.4948571428571429, "train_eval_precision": 0.4949971412235563,
"train_eval_recall": 0.999134448932487, "train_eval_f1": 0.662014911106863,
"train_eval_runtime": 15.3839, "train_eval_samples_per_second": 455.02,
"train_eval_steps_per_second": 3.575, "train_epoch": 1.0, "val_eval_loss":
1.7011138200759888, "val_eval_accuracy": 0.483, "val_eval_precision":
0.4824824824824825, "val_eval_recall": 1.0, "val_eval_f1": 0.6509115462525321,
"val_eval_runtime": 3.5933, "val_eval_samples_per_second": 278.294,
"val_eval_steps_per_second": 2.226, "val_epoch": 1.0, "test_eval_loss":
1.6855976581573486, "test_eval_accuracy": 0.48, "test_eval_precision":
0.48096192384769537, "test_eval_recall": 0.995850622406639, "test_eval_f1":
0.6486486486486487, "test_eval_runtime": 3.171, "test_eval_samples_per_second":
315.36, "test_eval_steps_per_second": 2.523, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:26:35.437831-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-large", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_dep_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.27.attn_norm.weight", "model.layers.27.attn.Wqkv.weight",
"model.layers.27.attn.Wo.weight", "model.layers.27.mlp_norm.weight",
"model.layers.27.mlp.Wi.weight", "model.layers.27.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 395833346, "model_trainable_params": 13309954,

```



```

"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-large",
"_commit_hash": "45bb4654a4d5aaff24dd11d4781fa46d39bf8c13",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 1024, "intermediate_size": 2624,
"num_hidden_layers": 28, "num_attention_heads": 16, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.7019022107124329, "train_eval_accuracy":
0.5201428571428571, "train_eval_precision": 0.5231101511879049,
"train_eval_recall": 0.3493941142527409, "train_eval_f1": 0.4189586576716831,
"train_eval_runtime": 39.3477, "train_eval_samples_per_second": 177.901,
"train_eval_steps_per_second": 1.398, "train_epoch": 1.0, "val_eval_loss":
0.6946681141853333, "val_eval_accuracy": 0.536, "val_eval_precision":
0.5272727272727272, "val_eval_recall": 0.36099585062240663, "val_eval_f1":
0.42857142857142855, "val_eval_runtime": 6.7532, "val_eval_samples_per_second":
148.079, "val_eval_steps_per_second": 1.185, "val_epoch": 1.0, "test_eval_loss":
0.7049111723899841, "test_eval_accuracy": 0.51, "test_eval_precision":
0.48830409356725146, "test_eval_recall": 0.34647302904564314, "test_eval_f1":

```

```

0.4053398058252427, "test_eval_runtime": 6.72, "test_eval_samples_per_second":
148.81, "test_eval_steps_per_second": 1.19, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:27:47.793263-07:00", "experiment_meta":
{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_dep_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",
"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",
"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":
"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":

```

```

null, "attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":
"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":
12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,
"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.6925950050354004, "train_eval_accuracy": 0.516,
"train_eval_precision": 0.5081047381546134, "train_eval_recall":
0.7054241200230814, "train_eval_f1": 0.5907223967141821, "train_eval_runtime":
14.7371, "train_eval_samples_per_second": 474.99, "train_eval_steps_per_second":
3.732, "train_epoch": 1.0, "val_eval_loss": 0.6941652894020081,
"val_eval_accuracy": 0.505, "val_eval_precision": 0.4906204906204906,
"val_eval_recall": 0.7053941908713693, "val_eval_f1": 0.5787234042553191,
"val_eval_runtime": 3.5136, "val_eval_samples_per_second": 284.607,
"val_eval_steps_per_second": 2.277, "val_epoch": 1.0, "test_eval_loss":
0.6938285231590271, "test_eval_accuracy": 0.488, "test_eval_precision":
0.4790502793296089, "test_eval_recall": 0.7116182572614108, "test_eval_f1":
0.5726210350584308, "test_eval_runtime": 3.1478, "test_eval_samples_per_second":
317.68, "test_eval_steps_per_second": 2.541, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:29:08.323398-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,

```

```

"task_specific_params": null, "problem_type": "single_label_classification",
"name_or_path": "bert-base-cased", "commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "attn_implementation_internal":
"sdpa", "attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6935000419616699, "train_eval_accuracy": 0.5076923076923077,
"train_eval_precision": 0.49787955894826125, "train_eval_recall":
0.9244094488188976, "train_eval_f1": 0.6471885336273429, "train_eval_runtime":
6.6479, "train_eval_samples_per_second": 195.55, "train_eval_steps_per_second":
1.655, "train_epoch": 1.0, "val_eval_loss": 0.7032004594802856,
"val_eval_accuracy": 0.468, "val_eval_precision": 0.4458874458874459,
"val_eval_recall": 0.9537037037037037, "val_eval_f1": 0.6076696165191741,
"val_eval_runtime": 6.0412, "val_eval_samples_per_second": 41.383,
"val_eval_steps_per_second": 0.331, "val_epoch": 1.0, "test_eval_loss":
0.69649338722229, "test_eval_accuracy": 0.496, "test_eval_precision":
0.5021645021645021, "test_eval_recall": 0.9133858267716536, "test_eval_f1":
0.6480446927374302, "test_eval_runtime": 5.1575, "test_eval_samples_per_second":
48.473, "test_eval_steps_per_second": 0.388, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:30:09.058297-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,

```

```

"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.693675696849823,
"train_eval_accuracy": 0.48846153846153845, "train_eval_precision":
0.48846153846153845, "train_eval_recall": 1.0, "train_eval_f1":
0.6563307493540051, "train_eval_runtime": 6.643,
"train_eval_samples_per_second": 195.694, "train_eval_steps_per_second": 1.656,
"train_epoch": 1.0, "val_eval_loss": 0.695859968662262, "val_eval_accuracy":
0.432, "val_eval_precision": 0.432, "val_eval_recall": 1.0, "val_eval_f1":
0.6033519553072626, "val_eval_runtime": 20.298, "val_eval_samples_per_second":
12.316, "val_eval_steps_per_second": 0.099, "val_epoch": 1.0, "test_eval_loss":
0.6926419734954834, "test_eval_accuracy": 0.508, "test_eval_precision": 0.508,
"test_eval_recall": 1.0, "test_eval_f1": 0.6737400530503979,
"test_eval_runtime": 5.1457, "test_eval_samples_per_second": 48.584,
"test_eval_steps_per_second": 0.389, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:32:12.987826-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,

```

```

"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaef65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidirectional", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.7076988220214844,
"train_eval_accuracy": 0.48, "train_eval_precision": 0.48301574150787074,
"train_eval_recall": 0.9181102362204724, "train_eval_f1": 0.6330076004343105,
"train_eval_runtime": 11.2415, "train_eval_samples_per_second": 115.642,
"train_eval_steps_per_second": 0.979, "train_epoch": 1.0, "val_eval_loss":
0.7160208821296692, "val_eval_accuracy": 0.42, "val_eval_precision":
0.4219409282700422, "val_eval_recall": 0.9259259259259259, "val_eval_f1":
0.5797101449275363, "val_eval_runtime": 7.2469, "val_eval_samples_per_second":
34.498, "val_eval_steps_per_second": 0.276, "val_epoch": 1.0, "test_eval_loss":
0.703961968421936, "test_eval_accuracy": 0.5, "test_eval_precision":
0.5042372881355932, "test_eval_recall": 0.937007874015748, "test_eval_f1":
0.6556473829201102, "test_eval_runtime": 6.621, "test_eval_samples_per_second":
37.759, "test_eval_steps_per_second": 0.302, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:33:20.607762-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":

```

```
[
  "roberta.encoder.layer.23.attention.self.query.weight",
  "roberta.encoder.layer.23.attention.self.query.bias",
  "roberta.encoder.layer.23.attention.self.key.weight",
  "roberta.encoder.layer.23.attention.self.key.bias",
  "roberta.encoder.layer.23.attention.self.value.weight",
  "roberta.encoder.layer.23.attention.self.value.bias",
  "roberta.encoder.layer.23.attention.output.dense.weight",
  "roberta.encoder.layer.23.attention.output.dense.bias",
  "roberta.encoder.layer.23.attention.output.LayerNorm.weight",
  "roberta.encoder.layer.23.attention.output.LayerNorm.bias",
  "roberta.encoder.layer.23.intermediate.dense.weight",
  "roberta.encoder.layer.23.intermediate.dense.bias",
  "roberta.encoder.layer.23.output.dense.weight",
  "roberta.encoder.layer.23.output.dense.bias",
  "roberta.encoder.layer.23.output.LayerNorm.weight",
  "roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
  "classifier.dense.bias", "classifier.out_proj.weight",
  "classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
  "model_trainable_params": 13647874, "pooler_activation_function": "N/A",
  "config_attributes": {"return_dict": true, "output_hidden_states": false,
  "output_attentions": false, "torchscript": false, "torch_dtype": "float32",
  "use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
  "tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
  "add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
  "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
  "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
  "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
  "bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
  "return_dict_in_generate": false, "forced_bos_token_id": null,
  "forced_eos_token_id": null, "remove_invalid_values": false,
  "exponential_decay_length_penalty": null, "suppress_tokens": null,
  "begin_suppress_tokens": null, "architectures":
  ["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
  "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
  "tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
  "eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
  "task_specific_params": null, "problem_type": "single_label_classification",
  "_name_or_path": "roberta-large", "_commit_hash":
  "722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
  "sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
  "model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
  "num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
  "intermediate_size": 4096, "hidden_dropout_prob": 0.1,
  "attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
  "type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
  "position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
```

```

null}}, "run_metrics": {"train_eval_loss": 0.6939926147460938,
"train_eval_accuracy": 0.48846153846153845, "train_eval_precision":
0.48846153846153845, "train_eval_recall": 1.0, "train_eval_f1":
0.6563307493540051, "train_eval_runtime": 11.2362,
"train_eval_samples_per_second": 115.697, "train_eval_steps_per_second": 0.979,
"train_epoch": 1.0, "val_eval_loss": 0.699723482131958, "val_eval_accuracy":
0.432, "val_eval_precision": 0.432, "val_eval_recall": 1.0, "val_eval_f1":
0.6033519553072626, "val_eval_runtime": 6.6693, "val_eval_samples_per_second":
37.485, "val_eval_steps_per_second": 0.3, "val_epoch": 1.0, "test_eval_loss":
0.692276120185852, "test_eval_accuracy": 0.508, "test_eval_precision": 0.508,
"test_eval_recall": 1.0, "test_eval_f1": 0.6737400530503979,
"test_eval_runtime": 6.0657, "test_eval_samples_per_second": 41.215,
"test_eval_steps_per_second": 0.33, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:34:28.312043-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_morph_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":

```



```

"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 2.163693904876709, "train_eval_accuracy":
0.48846153846153845, "train_eval_precision": 0.48846153846153845,
"train_eval_recall": 1.0, "train_eval_f1": 0.6563307493540051,
"train_eval_runtime": 7.9221, "train_eval_samples_per_second": 164.098,
"train_eval_steps_per_second": 1.389, "train_epoch": 1.0, "val_eval_loss":
2.3540005683898926, "val_eval_accuracy": 0.432, "val_eval_precision": 0.432,
"val_eval_recall": 1.0, "val_eval_f1": 0.6033519553072626, "val_eval_runtime":
6.0789, "val_eval_samples_per_second": 41.126, "val_eval_steps_per_second":
0.329, "val_epoch": 1.0, "test_eval_loss": 2.088582992553711,
"test_eval_accuracy": 0.508, "test_eval_precision": 0.508, "test_eval_recall":
1.0, "test_eval_f1": 0.6737400530503979, "test_eval_runtime": 5.473,
"test_eval_samples_per_second": 45.678, "test_eval_steps_per_second": 0.365,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:37:11.982255-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_dep_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},

```

```

"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdp", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7213355302810669, "train_eval_accuracy": 0.5115384615384615,
"train_eval_precision": 0.0, "train_eval_recall": 0.0, "train_eval_f1": 0.0,
"train_eval_runtime": 7.1615, "train_eval_samples_per_second": 181.526,
"train_eval_steps_per_second": 1.536, "train_epoch": 1.0, "val_eval_loss":
0.6946187615394592, "val_eval_accuracy": 0.568, "val_eval_precision": 0.0,
"val_eval_recall": 0.0, "val_eval_f1": 0.0, "val_eval_runtime": 5.3815,
"val_eval_samples_per_second": 46.455, "val_eval_steps_per_second": 0.372,
"val_epoch": 1.0, "test_eval_loss": 0.7301874756813049, "test_eval_accuracy":
0.492, "test_eval_precision": 0.0, "test_eval_recall": 0.0, "test_eval_f1": 0.0,
"test_eval_runtime": 5.2566, "test_eval_samples_per_second": 47.559,
"test_eval_steps_per_second": 0.38, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:38:03.678886-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_dep_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,

```

```

"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6942429542541504,
"train_eval_accuracy": 0.48846153846153845, "train_eval_precision":
0.48846153846153845, "train_eval_recall": 1.0, "train_eval_f1":
0.6563307493540051, "train_eval_runtime": 6.7781,
"train_eval_samples_per_second": 191.795, "train_eval_steps_per_second": 1.623,
"train_epoch": 1.0, "val_eval_loss": 0.6966016888618469, "val_eval_accuracy":
0.432, "val_eval_precision": 0.432, "val_eval_recall": 1.0, "val_eval_f1":
0.6033519553072626, "val_eval_runtime": 5.5199, "val_eval_samples_per_second":
45.291, "val_eval_steps_per_second": 0.362, "val_epoch": 1.0, "test_eval_loss":
0.6929192543029785, "test_eval_accuracy": 0.508, "test_eval_precision": 0.508,
"test_eval_recall": 1.0, "test_eval_f1": 0.6737400530503979,
"test_eval_runtime": 5.7039, "test_eval_samples_per_second": 43.83,
"test_eval_steps_per_second": 0.351, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:40:42.885286-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"snc_dep_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,

```

```

"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaef65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6973493099212646,
"train_eval_accuracy": 0.5307692307692308, "train_eval_precision":
0.5654450261780105, "train_eval_recall": 0.1700787401574803, "train_eval_f1":
0.26150121065375304, "train_eval_runtime": 10.6663,
"train_eval_samples_per_second": 121.88, "train_eval_steps_per_second": 1.031,
"train_epoch": 1.0, "val_eval_loss": 0.683283269405365, "val_eval_accuracy":
0.568, "val_eval_precision": 0.5, "val_eval_recall": 0.12037037037037036,
"val_eval_f1": 0.19402985074626866, "val_eval_runtime": 6.0207,
"val_eval_samples_per_second": 41.524, "val_eval_steps_per_second": 0.332,
"val_epoch": 1.0, "test_eval_loss": 0.7178395986557007, "test_eval_accuracy":
0.476, "test_eval_precision": 0.4411764705882353, "test_eval_recall":
0.11811023622047244, "test_eval_f1": 0.18633540372670807, "test_eval_runtime":
6.0972, "test_eval_samples_per_second": 41.002, "test_eval_steps_per_second":
0.328, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:41:51.325884-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":

```

```

"snc_dep_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",
"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,

```

```

"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6938353776931763,
"train_eval_accuracy": 0.48846153846153845, "train_eval_precision":
0.48846153846153845, "train_eval_recall": 1.0, "train_eval_f1":
0.6563307493540051, "train_eval_runtime": 11.2236,
"train_eval_samples_per_second": 115.827, "train_eval_steps_per_second": 0.98,
"train_epoch": 1.0, "val_eval_loss": 0.6985218524932861, "val_eval_accuracy":
0.432, "val_eval_precision": 0.432, "val_eval_recall": 1.0, "val_eval_f1":
0.6033519553072626, "val_eval_runtime": 6.1714, "val_eval_samples_per_second":
40.509, "val_eval_steps_per_second": 0.324, "val_epoch": 1.0, "test_eval_loss":
0.6930931806564331, "test_eval_accuracy": 0.508, "test_eval_precision": 0.508,
"test_eval_recall": 1.0, "test_eval_f1": 0.6737400530503979,
"test_eval_runtime": 6.5423, "test_eval_samples_per_second": 38.213,
"test_eval_steps_per_second": 0.306, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:44:21.628179-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_complexity_value", "y_col": "binary_complexity",
"layers_to_unfreeze": ["bert.encoder.layer.11.", "bert.pooler.",
"classifier."]}, "model_details": {"model_total_params": 108311810,
"model_trainable_params": 7680002, "pooler_activation_function": "Tanh",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,

```

```

"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6925702691078186, "train_eval_accuracy": 0.5077142857142857,
"train_eval_precision": 0.501929012345679, "train_eval_recall":
0.7507212925562609, "train_eval_f1": 0.6016184971098266, "train_eval_runtime":
15.1955, "train_eval_samples_per_second": 460.663,
"train_eval_steps_per_second": 3.619, "train_epoch": 1.0, "val_eval_loss":
0.6934520602226257, "val_eval_accuracy": 0.495, "val_eval_precision":
0.4844384303112314, "val_eval_recall": 0.7427385892116183, "val_eval_f1":
0.5864045864045864, "val_eval_runtime": 6.322, "val_eval_samples_per_second":
158.177, "val_eval_steps_per_second": 1.265, "val_epoch": 1.0, "test_eval_loss":
0.6921616196632385, "test_eval_accuracy": 0.522, "test_eval_precision":
0.5027548209366391, "test_eval_recall": 0.7572614107883817, "test_eval_f1":
0.6043046357615894, "test_eval_runtime": 6.7204, "test_eval_samples_per_second":
148.801, "test_eval_steps_per_second": 1.19, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:45:32.068713-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_morph_complexity_value", "y_col": "binary_complexity",
"layers_to_unfreeze": ["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":

```

```

1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6931387186050415,
"train_eval_accuracy": 0.501, "train_eval_precision": 0.4786053882725832,
"train_eval_recall": 0.08713214079630698, "train_eval_f1": 0.14742494508176715,
"train_eval_runtime": 15.4904, "train_eval_samples_per_second": 451.894,
"train_eval_steps_per_second": 3.551, "train_epoch": 1.0, "val_eval_loss":
0.692787766456604, "val_eval_accuracy": 0.519, "val_eval_precision":
0.5058823529411764, "val_eval_recall": 0.08921161825726141, "val_eval_f1":
0.15167548500881833, "val_eval_runtime": 7.0196, "val_eval_samples_per_second":
142.458, "val_eval_steps_per_second": 1.14, "val_epoch": 1.0, "test_eval_loss":
0.6928636431694031, "test_eval_accuracy": 0.525, "test_eval_precision":
0.5384615384615384, "test_eval_recall": 0.1016597510373444, "test_eval_f1":
0.17102966841186737, "test_eval_runtime": 6.2189,
"test_eval_samples_per_second": 160.801, "test_eval_steps_per_second": 1.286,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:47:13.702804-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_pos_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":

```



```

50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6946821212768555, "train_eval_accuracy": 0.49014285714285716,
"train_eval_precision": 0.48345647285576615, "train_eval_recall":
0.4342181188690133, "train_eval_f1": 0.45751633986928103, "train_eval_runtime":
11.2374, "train_eval_samples_per_second": 622.923,
"train_eval_steps_per_second": 4.894, "train_epoch": 1.0, "val_eval_loss":
0.6940599679946899, "val_eval_accuracy": 0.513, "val_eval_precision":
0.494199535962877, "val_eval_recall": 0.44190871369294604, "val_eval_f1":
0.4665936473165389, "val_eval_runtime": 2.5962, "val_eval_samples_per_second":
385.184, "val_eval_steps_per_second": 3.081, "val_epoch": 1.0, "test_eval_loss":
0.6929909586906433, "test_eval_accuracy": 0.519, "test_eval_precision":
0.5011086474501109, "test_eval_recall": 0.46887966804979253, "test_eval_f1":
0.4844587352625938, "test_eval_runtime": 2.5932, "test_eval_samples_per_second":
385.621, "test_eval_steps_per_second": 3.085, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:47:57.583475-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_pos_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",

```

```

"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6932085752487183,
"train_eval_accuracy": 0.5028571428571429, "train_eval_precision":
0.4351851851851852, "train_eval_recall": 0.013560300057703404, "train_eval_f1":
0.02630106323447118, "train_eval_runtime": 11.7655,
"train_eval_samples_per_second": 594.958, "train_eval_steps_per_second": 4.675,
"train_epoch": 1.0, "val_eval_loss": 0.6928229928016663, "val_eval_accuracy":
0.518, "val_eval_precision": 0.5, "val_eval_recall": 0.01867219917012448,
"val_eval_f1": 0.036, "val_eval_runtime": 2.4822, "val_eval_samples_per_second":
402.871, "val_eval_steps_per_second": 3.223, "val_epoch": 1.0, "test_eval_loss":

```

```

0.6929774284362793, "test_eval_accuracy": 0.521, "test_eval_precision":
0.5555555555555556, "test_eval_recall": 0.03112033195020747, "test_eval_f1":
0.05893909626719057, "test_eval_runtime": 2.4672,
"test_eval_samples_per_second": 405.325, "test_eval_steps_per_second": 3.243,
"test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:50:55.138111-07:00", "experiment_meta":
{"model_name": "bert-large-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_pos_alt", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.23.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 333581314, "model_trainable_params": 13647874,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaeff65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6915509104728699,
"train_eval_accuracy": 0.5314285714285715, "train_eval_precision":
0.5203056768558952, "train_eval_recall": 0.687536064627813, "train_eval_f1":
0.5923440218742232, "train_eval_runtime": 33.2325,
"train_eval_samples_per_second": 210.637, "train_eval_steps_per_second": 1.655,

```

```

"train_epoch": 1.0, "val_eval_loss": 0.6909124851226807, "val_eval_accuracy":
0.535, "val_eval_precision": 0.512630014858841, "val_eval_recall":
0.7157676348547718, "val_eval_f1": 0.5974025974025974, "val_eval_runtime":
5.9766, "val_eval_samples_per_second": 167.319, "val_eval_steps_per_second":
1.339, "val_epoch": 1.0, "test_eval_loss": 0.6918430328369141,
"test_eval_accuracy": 0.526, "test_eval_precision": 0.5059523809523809,
"test_eval_recall": 0.7053941908713693, "test_eval_f1": 0.5892547660311959,
"test_eval_runtime": 5.8682, "test_eval_samples_per_second": 170.41,
"test_eval_steps_per_second": 1.363, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:52:08.371010-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.6928351521492004, "train_eval_accuracy": 0.5138571428571429,
"train_eval_precision": 0.5054964229628337, "train_eval_recall":
0.8358338141950375, "train_eval_f1": 0.6299880395781233, "train_eval_runtime":

```

```

11.56, "train_eval_samples_per_second": 605.537, "train_eval_steps_per_second":
4.758, "train_epoch": 1.0, "val_eval_loss": 0.6933068037033081,
"val_eval_accuracy": 0.516, "val_eval_precision": 0.4987714987714988,
"val_eval_recall": 0.8423236514522822, "val_eval_f1": 0.6265432098765432,
"val_eval_runtime": 3.2822, "val_eval_samples_per_second": 304.675,
"val_eval_steps_per_second": 2.437, "val_epoch": 1.0, "test_eval_loss":
0.6943942904472351, "test_eval_accuracy": 0.5, "test_eval_precision":
0.48933649289099523, "test_eval_recall": 0.8568464730290456, "test_eval_f1":
0.6229260935143288, "test_eval_runtime": 3.8898, "test_eval_samples_per_second":
257.082, "test_eval_steps_per_second": 2.057, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:52:54.064322-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",
"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":

```

```
[
  "RobertaForSequenceClassification",
  "finetuning_task": null,
  "id2label": {"0": "LABEL_0", "1": "LABEL_1"},
  "label2id": {"LABEL_0": 0, "LABEL_1": 1},
  "tokenizer_class": null,
  "prefix": null,
  "bos_token_id": 0,
  "pad_token_id": 1,
  "eos_token_id": 2,
  "sep_token_id": null,
  "decoder_start_token_id": null,
  "task_specific_params": null,
  "problem_type": "single_label_classification",
  "_name_or_path": "roberta-base",
  "_commit_hash": "e2da8e2f811d1448a5b465c236feacd80ffb7b",
  "_attn_implementation_internal": "sdpa",
  "_attn_implementation_autoset": false,
  "transformers_version": null,
  "model_type": "roberta",
  "vocab_size": 50265,
  "hidden_size": 768,
  "num_hidden_layers": 12,
  "num_attention_heads": 12,
  "hidden_act": "gelu",
  "intermediate_size": 3072,
  "hidden_dropout_prob": 0.1,
  "attention_probs_dropout_prob": 0.1,
  "max_position_embeddings": 514,
  "type_vocab_size": 1,
  "initializer_range": 0.02,
  "layer_norm_eps": 1e-05,
  "position_embedding_type": "absolute",
  "use_cache": true,
  "classifier_dropout": null,
  "run_metrics": {
    "train_eval_loss": 0.6931005120277405,
    "train_eval_accuracy": 0.5052857142857143,
    "train_eval_precision": 0.5089820359281437,
    "train_eval_recall": 0.02452394691286786,
    "train_eval_f1": 0.04679328378750344,
    "train_eval_runtime": 12.1462,
    "train_eval_samples_per_second": 576.311,
    "train_eval_steps_per_second": 4.528,
    "train_epoch": 1.0,
    "val_eval_loss": 0.6926701664924622,
    "val_eval_accuracy": 0.517,
    "val_eval_precision": 0.4827586206896552,
    "val_eval_recall": 0.029045643153526972,
    "val_eval_f1": 0.0547945205479452,
    "val_eval_runtime": 2.9994,
    "val_eval_samples_per_second": 333.398,
    "val_eval_steps_per_second": 2.667,
    "val_epoch": 1.0,
    "test_eval_loss": 0.6928783059120178,
    "test_eval_accuracy": 0.518,
    "test_eval_precision": 0.5,
    "test_eval_recall": 0.024896265560165973,
    "test_eval_f1": 0.04743083003952569,
    "test_eval_runtime": 2.8514,
    "test_eval_samples_per_second": 350.702,
    "test_eval_steps_per_second": 2.806,
    "test_epoch": 1.0
  }
},
{
  "timestamp": "2025-04-11T14:56:00.272646-07:00",
  "experiment_meta": {
    "model_name": "bert-large-cased",
    "learning_rate": 5e-06,
    "epochs": 1,
    "batch_size": 128,
    "weight_decay": 0.5,
    "x_task": "single",
    "x_col": "snc_pos_seq",
    "y_col": "binary_complexity",
    "layers_to_unfreeze": ["bert.encoder.layer.23.", "bert.pooler.", "classifier."],
    "model_details": {
      "model_total_params": 333581314,
      "model_trainable_params": 13647874,
      "pooler_activation_function": "Tanh",
      "config_attributes": {
        "return_dict": true,
        "output_hidden_states": false,
        "output_attentions": false,
        "torchscript": false,
        "torch_dtype": "float32",
        "use_bfloat16": false,
        "tf_legacy_loss": false,
        "pruned_heads": {},
        "tie_word_embeddings": true,
        "chunk_size_feed_forward": 0,
        "is_encoder_decoder": false,
        "is_decoder": false,
        "cross_attention_hidden_size": null,
        "add_cross_attention": false,
        "tie_encoder_decoder": false,
        "max_length": 20,
        "min_length": 0,
        "do_sample": false,
        "early_stopping": false,
        "num_beams": 1,
        "num_beam_groups": 1,
        "diversity_penalty": 0.0,
        "temperature": 1.0,
        "top_k": 50,
        "top_p": 1.0,
        "typical_p": 1.0,
        "repetition_penalty": 1.0,
        "length_penalty": 1.0,
        "no_repeat_ngram_size": 0,
        "encoder_no_repeat_ngram_size": 0,
        "bad_words_ids": null,
        "num_return_sequences": 1,
        "output_scores": false,
        "return_dict_in_generate": false,
        "forced_bos_token_id": null,
        "forced_eos_token_id": null,
        "remove_invalid_values": false,
        "exponential_decay_length_penalty": null,
        "suppress_tokens": null,

```

```

"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-large-cased", "_commit_hash":
"06fa25dc1b6e034e91607fd67c247051ccaeff65", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "directionality": "bidi", "gradient_checkpointing": false,
"model_type": "bert", "pooler_fc_size": 768, "pooler_num_attention_heads": 12,
"pooler_num_fc_layers": 3, "pooler_size_per_head": 128, "pooler_type":
"first_token_transform", "vocab_size": 28996, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 512,
"type_vocab_size": 2, "initializer_range": 0.02, "layer_norm_eps": 1e-12,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6946611404418945,
"train_eval_accuracy": 0.5092857142857142, "train_eval_precision":
0.5060147458284827, "train_eval_recall": 0.3762261973456434, "train_eval_f1":
0.43157372166142643, "train_eval_runtime": 33.5525,
"train_eval_samples_per_second": 208.628, "train_eval_steps_per_second": 1.639,
"train_epoch": 1.0, "val_eval_loss": 0.696256697177887, "val_eval_accuracy":
0.514, "val_eval_precision": 0.4948186528497409, "val_eval_recall":
0.3962655601659751, "val_eval_f1": 0.4400921658986175, "val_eval_runtime":
5.7987, "val_eval_samples_per_second": 172.451, "val_eval_steps_per_second":
1.38, "val_epoch": 1.0, "test_eval_loss": 0.6938513517379761,
"test_eval_accuracy": 0.503, "test_eval_precision": 0.47989276139410186,
"test_eval_recall": 0.37136929460580914, "test_eval_f1": 0.41871345029239765,
"test_eval_runtime": 5.8466, "test_eval_samples_per_second": 171.04,
"test_eval_steps_per_second": 1.368, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:57:52.261163-07:00", "experiment_meta":
{"model_name": "roberta-large", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.23.attention.self.query.weight",
"roberta.encoder.layer.23.attention.self.query.bias",
"roberta.encoder.layer.23.attention.self.key.weight",
"roberta.encoder.layer.23.attention.self.key.bias",
"roberta.encoder.layer.23.attention.self.value.weight",
"roberta.encoder.layer.23.attention.self.value.bias",
"roberta.encoder.layer.23.attention.output.dense.weight",
"roberta.encoder.layer.23.attention.output.dense.bias",
"roberta.encoder.layer.23.attention.output.LayerNorm.weight",
"roberta.encoder.layer.23.attention.output.LayerNorm.bias",
"roberta.encoder.layer.23.intermediate.dense.weight",
"roberta.encoder.layer.23.intermediate.dense.bias",

```

```

"roberta.encoder.layer.23.output.dense.weight",
"roberta.encoder.layer.23.output.dense.bias",
"roberta.encoder.layer.23.output.LayerNorm.weight",
"roberta.encoder.layer.23.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 355361794,
"model_trainable_params": 13647874, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-large", "_commit_hash":
"722cf37b1afa9454edce342e7895e588b6ff1d59", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 1024,
"num_hidden_layers": 24, "num_attention_heads": 16, "hidden_act": "gelu",
"intermediate_size": 4096, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6924602389335632,
"train_eval_accuracy": 0.517, "train_eval_precision": 0.5086434817978442,
"train_eval_recall": 0.7215810732833238, "train_eval_f1": 0.5966837647620183,
"train_eval_runtime": 33.3444, "train_eval_samples_per_second": 209.93,
"train_eval_steps_per_second": 1.649, "train_epoch": 1.0, "val_eval_loss":
0.6925870180130005, "val_eval_accuracy": 0.534, "val_eval_precision":
0.511049723756906, "val_eval_recall": 0.7676348547717843, "val_eval_f1":
0.6135986733001658, "val_eval_runtime": 5.9644, "val_eval_samples_per_second":
167.661, "val_eval_steps_per_second": 1.341, "val_epoch": 1.0, "test_eval_loss":
0.6931580901145935, "test_eval_accuracy": 0.511, "test_eval_precision":
0.4948905109489051, "test_eval_recall": 0.7033195020746889, "test_eval_f1":
0.5809768637532133, "test_eval_runtime": 5.8298, "test_eval_samples_per_second":

```



```

171.533, "test_eval_steps_per_second": 1.372, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T14:59:13.682519-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":

```

```

{"train_eval_loss": 1.6881275177001953, "train_eval_accuracy": 0.496,
"train_eval_precision": 0.4955536431440046, "train_eval_recall":
0.9968263127524524, "train_eval_f1": 0.6620042153669286, "train_eval_runtime":
16.2713, "train_eval_samples_per_second": 430.206,
"train_eval_steps_per_second": 3.38, "train_epoch": 1.0, "val_eval_loss":
1.7097792625427246, "val_eval_accuracy": 0.484, "val_eval_precision":
0.48296593186372744, "val_eval_recall": 1.0, "val_eval_f1": 0.6513513513513514,
"val_eval_runtime": 3.1147, "val_eval_samples_per_second": 321.056,
"val_eval_steps_per_second": 2.568, "val_epoch": 1.0, "test_eval_loss":
1.719785213470459, "test_eval_accuracy": 0.481, "test_eval_precision":
0.48148148148148145, "test_eval_recall": 0.9979253112033195, "test_eval_f1":
0.6495611073598919, "test_eval_runtime": 3.6075, "test_eval_samples_per_second":
277.201, "test_eval_steps_per_second": 2.218, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T15:01:34.715598-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-large", "learning_rate": 5e-06, "epochs":
1, "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.27.attn_norm.weight", "model.layers.27.attn.Wqkv.weight",
"model.layers.27.attn.Wo.weight", "model.layers.27.mlp_norm.weight",
"model.layers.27.mlp.Wi.weight", "model.layers.27.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 395833346, "model_trainable_params": 13309954,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-large",
"_commit_hash": "45bb4654a4d5aaff24dd11d4781fa46d39bf8c13",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":

```

```

"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 1024, "intermediate_size": 2624,
"num_hidden_layers": 28, "num_attention_heads": 16, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.6943276524543762, "train_eval_accuracy":
0.5274285714285715, "train_eval_precision": 0.523235294117647,
"train_eval_recall": 0.513271783035199, "train_eval_f1": 0.518205651034081,
"train_eval_runtime": 39.4979, "train_eval_samples_per_second": 177.225,
"train_eval_steps_per_second": 1.392, "train_epoch": 1.0, "val_eval_loss":
0.6854519844055176, "val_eval_accuracy": 0.551, "val_eval_precision":
0.534446764091858, "val_eval_recall": 0.5311203319502075, "val_eval_f1":
0.5327783558792925, "val_eval_runtime": 7.0263, "val_eval_samples_per_second":
142.323, "val_eval_steps_per_second": 1.139, "val_epoch": 1.0, "test_eval_loss":
0.6978042721748352, "test_eval_accuracy": 0.542, "test_eval_precision":
0.5253164556962026, "test_eval_recall": 0.516597510373444, "test_eval_f1":
0.5209205020920502, "test_eval_runtime": 6.7568, "test_eval_samples_per_second":
148.0, "test_eval_steps_per_second": 1.184, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T15:02:45.629153-07:00", "experiment_meta":
{"model_name": "microsoft/deberta-v3-base", "learning_rate": 5e-06, "epochs": 1,
"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["deberta.encoder.layer.11.attention.self.query_proj.weight",
"deberta.encoder.layer.11.attention.self.query_proj.bias",
"deberta.encoder.layer.11.attention.self.key_proj.weight",
"deberta.encoder.layer.11.attention.self.key_proj.bias",
"deberta.encoder.layer.11.attention.self.value_proj.weight",
"deberta.encoder.layer.11.attention.self.value_proj.bias",
"deberta.encoder.layer.11.attention.output.dense.weight",
"deberta.encoder.layer.11.attention.output.dense.bias",
"deberta.encoder.layer.11.attention.output.LayerNorm.weight",
"deberta.encoder.layer.11.attention.output.LayerNorm.bias",
"deberta.encoder.layer.11.intermediate.dense.weight",
"deberta.encoder.layer.11.intermediate.dense.bias",
"deberta.encoder.layer.11.output.dense.weight",
"deberta.encoder.layer.11.output.dense.bias",
"deberta.encoder.layer.11.output.LayerNorm.weight",
"deberta.encoder.layer.11.output.LayerNorm.bias",
"deberta.encoder.rel_embeddings.weight", "deberta.encoder.LayerNorm.weight",
"deberta.encoder.LayerNorm.bias", "pooler.dense.weight", "pooler.dense.bias",
"classifier.weight", "classifier.bias"]}, "model_details":

```

```

{"model_total_params": 184423682, "model_trainable_params": 8074754,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["DebertaV2ForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": null, "_name_or_path":
"microsoft/deberta-v3-base", "_commit_hash":
"8ccc9b6f36199bec6961081d44eb72fb3f7353f3", "_attn_implementation_internal":
null, "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "deberta-v2", "position_buckets": 256, "norm_rel_ebd":
"layer_norm", "share_att_key": true, "hidden_size": 768, "num_hidden_layers":
12, "num_attention_heads": 12, "intermediate_size": 3072, "hidden_act": "gelu",
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 0, "initializer_range": 0.02,
"relative_attention": true, "max_relative_positions": -1,
"position_biased_input": false, "pos_att_type": ["p2c", "c2p"], "vocab_size":
128100, "layer_norm_eps": 1e-07, "pooler_hidden_size": 768, "pooler_dropout": 0,
"pooler_hidden_act": "gelu", "legacy": true}}, "run_metrics":
{"train_eval_loss": 0.6949706077575684, "train_eval_accuracy":
0.5005714285714286, "train_eval_precision": 0.49753208292201384,
"train_eval_recall": 0.8724754760530872, "train_eval_f1": 0.633696563285834,
"train_eval_runtime": 14.6619, "train_eval_samples_per_second": 477.428,
"train_eval_steps_per_second": 3.751, "train_epoch": 1.0, "val_eval_loss":
0.6953136324882507, "val_eval_accuracy": 0.504, "val_eval_precision":
0.4920273348519362, "val_eval_recall": 0.8962655601659751, "val_eval_f1":
0.6352941176470588, "val_eval_runtime": 3.0848, "val_eval_samples_per_second":
324.171, "val_eval_steps_per_second": 2.593, "val_epoch": 1.0, "test_eval_loss":
0.6948614716529846, "test_eval_accuracy": 0.489, "test_eval_precision":
0.48342857142857143, "test_eval_recall": 0.8775933609958506, "test_eval_f1":
0.6234340456890199, "test_eval_runtime": 3.2109, "test_eval_samples_per_second":
311.44, "test_eval_steps_per_second": 2.492, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T15:05:08.378028-07:00", "experiment_meta":
{"model_name": "xlnet/xlnet-base-cased", "learning_rate": 5e-06, "epochs": 1,

```

```

"batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["transformer.layer.11.rel_attn.q", "transformer.layer.11.rel_attn.k",
"transformer.layer.11.rel_attn.v", "transformer.layer.11.rel_attn.o",
"transformer.layer.11.rel_attn.r", "transformer.layer.11.rel_attn.r_r_bias",
"transformer.layer.11.rel_attn.r_s_bias",
"transformer.layer.11.rel_attn.r_w_bias",
"transformer.layer.11.rel_attn.seg_embed",
"transformer.layer.11.rel_attn.layer_norm.weight",
"transformer.layer.11.rel_attn.layer_norm.bias",
"transformer.layer.11.ff.layer_norm.weight",
"transformer.layer.11.ff.layer_norm.bias",
"transformer.layer.11.ff.layer_1.weight",
"transformer.layer.11.ff.layer_1.bias",
"transformer.layer.11.ff.layer_2.weight",
"transformer.layer.11.ff.layer_2.bias", "sequence_summary.summary.weight",
"sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
"model_details": {"model_total_params": 117310466, "model_trainable_params":
8270594, "pooler_activation_function": "N/A", "config_attributes":
{"vocab_size": 32000, "d_model": 768, "n_layer": 12, "n_head": 12, "d_head": 64,
"ff_activation": "gelu", "d_inner": 3072, "untie_r": true, "attn_type": "bi",
"initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1, "mem_len":
null, "reuse_len": null, "bi_data": false, "clamp_len": -1, "same_length":
false, "summary_type": "last", "summary_use_proj": true, "summary_activation":
"tanh", "summary_last_dropout": 0.1, "start_n_top": 5, "end_n_top": 5,
"bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2, "use_mems_eval": true,
"use_mems_train": false, "return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-base-cased",
"_commit_hash": "ceaa69c7bc5e512b5007106a7ccbb7daf24b2c79",

```

```

    "_attn_implementation_internal": null, "_attn_implementation_autoset": false,
    "transformers_version": null, "model_type": "xlnet"}}, "run_metrics":
    {"train_eval_loss": 0.6996774673461914, "train_eval_accuracy":
    0.5044285714285714, "train_eval_precision": 0.49910125823846613,
    "train_eval_recall": 0.240334679746105, "train_eval_f1": 0.32444011684518015,
    "train_eval_runtime": 18.0108, "train_eval_samples_per_second": 388.656,
    "train_eval_steps_per_second": 3.054, "train_epoch": 1.0, "val_eval_loss":
    0.6945834159851074, "val_eval_accuracy": 0.518, "val_eval_precision": 0.5,
    "val_eval_recall": 0.2551867219917012, "val_eval_f1": 0.33791208791208793,
    "val_eval_runtime": 5.21, "val_eval_samples_per_second": 191.939,
    "val_eval_steps_per_second": 1.536, "val_epoch": 1.0, "test_eval_loss":
    0.6950303912162781, "test_eval_accuracy": 0.525, "test_eval_precision":
    0.5155555555555555, "test_eval_recall": 0.24066390041493776, "test_eval_f1":
    0.32814710042432815, "test_eval_runtime": 3.6324,
    "test_eval_samples_per_second": 275.298, "test_eval_steps_per_second": 2.202,
    "test_epoch": 1.0}}
    {"timestamp": "2025-04-11T15:07:53.771109-07:00", "experiment_meta":
    {"model_name": "xlnet/xlnet-large-cased", "learning_rate": 5e-06, "epochs": 1,
    "batch_size": 128, "weight_decay": 0.5, "x_task": "single", "x_col":
    "snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
    ["transformer.layer.23.rel_attn.q", "transformer.layer.23.rel_attn.k",
    "transformer.layer.23.rel_attn.v", "transformer.layer.23.rel_attn.o",
    "transformer.layer.23.rel_attn.r", "transformer.layer.23.rel_attn.r_r_bias",
    "transformer.layer.23.rel_attn.r_s_bias",
    "transformer.layer.23.rel_attn.r_w_bias",
    "transformer.layer.23.rel_attn.seg_embed",
    "transformer.layer.23.rel_attn.layer_norm.weight",
    "transformer.layer.23.rel_attn.layer_norm.bias",
    "transformer.layer.23.ff.layer_norm.weight",
    "transformer.layer.23.ff.layer_norm.bias",
    "transformer.layer.23.ff.layer_1.weight",
    "transformer.layer.23.ff.layer_1.bias",
    "transformer.layer.23.ff.layer_2.weight",
    "transformer.layer.23.ff.layer_2.bias", "sequence_summary.summary.weight",
    "sequence_summary.summary.bias", "logits_proj.weight", "logits_proj.bias"]},
    "model_details": {"model_total_params": 361320450, "model_trainable_params":
    14697474, "pooler_activation_function": "N/A", "config_attributes":
    {"vocab_size": 32000, "d_model": 1024, "n_layer": 24, "n_head": 16, "d_head":
    64, "ff_activation": "gelu", "d_inner": 4096, "untie_r": true, "attn_type":
    "bi", "initializer_range": 0.02, "layer_norm_eps": 1e-12, "dropout": 0.1,
    "mem_len": null, "reuse_len": null, "bi_data": false, "clamp_len": -1,
    "same_length": false, "summary_type": "last", "summary_use_proj": true,
    "summary_activation": "tanh", "summary_last_dropout": 0.1, "start_n_top": 5,
    "end_n_top": 5, "bos_token_id": 1, "pad_token_id": 5, "eos_token_id": 2,
    "use_mems_eval": true, "use_mems_train": false, "return_dict": true,
    "output_hidden_states": false, "output_attentions": false, "torchscript": false,
    "torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
    "pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,

```

```

"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["XLNetForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "sep_token_id": null,
"decoder_start_token_id": null, "task_specific_params": {"text-generation":
{"do_sample": true, "max_length": 250}}, "problem_type":
"single_label_classification", "_name_or_path": "xlnet/xlnet-large-cased",
"_commit_hash": "37658b4f179eaf971d127d16bfbfd9ca676a93034",
"_attn_implementation_internal": null, "_attn_implementation_autoset": false,
"transformers_version": null, "model_type": "xlnet"}}}, "run_metrics":
{"train_eval_loss": 0.6902309656143188, "train_eval_accuracy":
0.5361428571428571, "train_eval_precision": 0.5316017316017316,
"train_eval_recall": 0.5314483554529718, "train_eval_f1": 0.5315250324628481,
"train_eval_runtime": 50.9469, "train_eval_samples_per_second": 137.398,
"train_eval_steps_per_second": 1.08, "train_epoch": 1.0, "val_eval_loss":
0.691400945186615, "val_eval_accuracy": 0.531, "val_eval_precision":
0.5139186295503212, "val_eval_recall": 0.4979253112033195, "val_eval_f1":
0.505795574288725, "val_eval_runtime": 8.3782, "val_eval_samples_per_second":
119.357, "val_eval_steps_per_second": 0.955, "val_epoch": 1.0, "test_eval_loss":
0.6909890174865723, "test_eval_accuracy": 0.529, "test_eval_precision":
0.5110220440881763, "test_eval_recall": 0.529045643153527, "test_eval_f1":
0.5198776758409785, "test_eval_runtime": 8.5255, "test_eval_samples_per_second":
117.296, "test_eval_steps_per_second": 0.938, "test_epoch": 1.0}}
{"timestamp": "2025-04-11T21:44:10.180778-07:00", "experiment_meta":
{"model_name": "bert-base-cased", "learning_rate": 5e-09, "epochs": 5,
"batch_size": 128, "weight_decay": 0.25, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["bert.encoder.layer.11.", "bert.pooler.", "classifier."]}, "model_details":
{"model_total_params": 108311810, "model_trainable_params": 7680002,
"pooler_activation_function": "Tanh", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":

```

```

1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["BertForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": null, "pad_token_id":
0, "eos_token_id": null, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "bert-base-cased", "_commit_hash":
"cd5ef92a9fb2f889e972770a36d4ed042daf221e", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version":
"4.6.0.dev0", "gradient_checkpointing": false, "model_type": "bert",
"vocab_size": 28996, "hidden_size": 768, "num_hidden_layers": 12,
"num_attention_heads": 12, "hidden_act": "gelu", "intermediate_size": 3072,
"hidden_dropout_prob": 0.1, "attention_probs_dropout_prob": 0.1,
"max_position_embeddings": 512, "type_vocab_size": 2, "initializer_range": 0.02,
"layer_norm_eps": 1e-12, "position_embedding_type": "absolute", "use_cache":
true, "classifier_dropout": null}}, "run_metrics": {"train_eval_loss":
0.7908573150634766, "train_eval_accuracy": 0.48846153846153845,
"train_eval_precision": 0.48846153846153845, "train_eval_recall": 1.0,
"train_eval_f1": 0.6563307493540051, "train_eval_runtime": 12.912,
"train_eval_samples_per_second": 100.682, "train_eval_steps_per_second": 0.852,
"train_epoch": 5.0, "val_eval_loss": 0.8314957022666931, "val_eval_accuracy":
0.432, "val_eval_precision": 0.432, "val_eval_recall": 1.0, "val_eval_f1":
0.6033519553072626, "val_eval_runtime": 5.5391, "val_eval_samples_per_second":
45.134, "val_eval_steps_per_second": 0.361, "val_epoch": 5.0, "test_eval_loss":
0.7790234088897705, "test_eval_accuracy": 0.508, "test_eval_precision": 0.508,
"test_eval_recall": 1.0, "test_eval_f1": 0.6737400530503979,
"test_eval_runtime": 5.5541, "test_eval_samples_per_second": 45.012,
"test_eval_steps_per_second": 0.36, "test_epoch": 5.0}}
{"timestamp": "2025-04-11T21:45:28.890059-07:00", "experiment_meta":
{"model_name": "roberta-base", "learning_rate": 5e-09, "epochs": 5,
"batch_size": 128, "weight_decay": 0.5, "x_task": "multi", "x_col":
"sentence_no_contractions", "y_col": "binary_complexity", "layers_to_unfreeze":
["roberta.encoder.layer.11.attention.self.query.weight",
"roberta.encoder.layer.11.attention.self.query.bias",
"roberta.encoder.layer.11.attention.self.key.weight",
"roberta.encoder.layer.11.attention.self.key.bias",
"roberta.encoder.layer.11.attention.self.value.weight",
"roberta.encoder.layer.11.attention.self.value.bias",
"roberta.encoder.layer.11.attention.output.dense.weight",
"roberta.encoder.layer.11.attention.output.dense.bias",
"roberta.encoder.layer.11.attention.output.LayerNorm.weight",
"roberta.encoder.layer.11.attention.output.LayerNorm.bias",
"roberta.encoder.layer.11.intermediate.dense.weight",

```



```

"roberta.encoder.layer.11.intermediate.dense.bias",
"roberta.encoder.layer.11.output.dense.weight",
"roberta.encoder.layer.11.output.dense.bias",
"roberta.encoder.layer.11.output.LayerNorm.weight",
"roberta.encoder.layer.11.output.LayerNorm.bias", "classifier.dense.weight",
"classifier.dense.bias", "classifier.out_proj.weight",
"classifier.out_proj.bias"]}, "model_details": {"model_total_params": 124647170,
"model_trainable_params": 7680002, "pooler_activation_function": "N/A",
"config_attributes": {"return_dict": true, "output_hidden_states": false,
"output_attentions": false, "torchscript": false, "torch_dtype": "float32",
"use_bfloat16": false, "tf_legacy_loss": false, "pruned_heads": {}},
"tie_word_embeddings": true, "chunk_size_feed_forward": 0, "is_encoder_decoder":
false, "is_decoder": false, "cross_attention_hidden_size": null,
"add_cross_attention": false, "tie_encoder_decoder": false, "max_length": 20,
"min_length": 0, "do_sample": false, "early_stopping": false, "num_beams": 1,
"num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k": 50,
"top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["RobertaForSequenceClassification"], "finetuning_task": null, "id2label": {"0":
"LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 0, "pad_token_id": 1,
"eos_token_id": 2, "sep_token_id": null, "decoder_start_token_id": null,
"task_specific_params": null, "problem_type": "single_label_classification",
"_name_or_path": "roberta-base", "_commit_hash":
"e2da8e2f811d1448a5b465c236feacd80ffbac7b", "_attn_implementation_internal":
"sdpa", "_attn_implementation_autoset": false, "transformers_version": null,
"model_type": "roberta", "vocab_size": 50265, "hidden_size": 768,
"num_hidden_layers": 12, "num_attention_heads": 12, "hidden_act": "gelu",
"intermediate_size": 3072, "hidden_dropout_prob": 0.1,
"attention_probs_dropout_prob": 0.1, "max_position_embeddings": 514,
"type_vocab_size": 1, "initializer_range": 0.02, "layer_norm_eps": 1e-05,
"position_embedding_type": "absolute", "use_cache": true, "classifier_dropout":
null}}, "run_metrics": {"train_eval_loss": 0.6939494013786316,
"train_eval_accuracy": 0.48846153846153845, "train_eval_precision":
0.48846153846153845, "train_eval_recall": 1.0, "train_eval_f1":
0.6563307493540051, "train_eval_runtime": 6.775,
"train_eval_samples_per_second": 191.881, "train_eval_steps_per_second": 1.624,
"train_epoch": 5.0, "val_eval_loss": 0.6972026228904724, "val_eval_accuracy":
0.432, "val_eval_precision": 0.432, "val_eval_recall": 1.0, "val_eval_f1":
0.6033519553072626, "val_eval_runtime": 5.4507, "val_eval_samples_per_second":
45.865, "val_eval_steps_per_second": 0.367, "val_epoch": 5.0, "test_eval_loss":
0.6924076080322266, "test_eval_accuracy": 0.508, "test_eval_precision": 0.508,
"test_eval_recall": 1.0, "test_eval_f1": 0.6737400530503979,

```

```

"test_eval_runtime": 5.2641, "test_eval_samples_per_second": 47.491,
"test_eval_steps_per_second": 0.38, "test_epoch": 5.0}}
{"timestamp": "2025-04-12T07:52:36.512068-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 5e-07, "epochs":
5, "batch_size": 1, "weight_decay": 0, "x_task": "single", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,

```

```

"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 1.03140127658844, "train_eval_accuracy": 0.5324285714285715,
"train_eval_precision": 0.5636123928806855, "train_eval_recall":
0.24668205424120024, "train_eval_f1": 0.3431667670078266, "train_eval_runtime":
211.3364, "train_eval_samples_per_second": 33.123,
"train_eval_steps_per_second": 33.123, "train_epoch": 5.0, "val_eval_loss":
0.9834219813346863, "val_eval_accuracy": 0.539, "val_eval_precision":
0.5488372093023256, "val_eval_recall": 0.24481327800829875, "val_eval_f1":
0.3385939741750359, "val_eval_runtime": 30.9148, "val_eval_samples_per_second":
32.347, "val_eval_steps_per_second": 32.347, "val_epoch": 5.0, "test_eval_loss":
1.0433307886123657, "test_eval_accuracy": 0.534, "test_eval_precision":
0.536036036036036, "test_eval_recall": 0.24688796680497926, "test_eval_f1":
0.3380681818181818, "test_eval_runtime": 30.896, "test_eval_samples_per_second":
32.367, "test_eval_steps_per_second": 32.367, "test_epoch": 5.0}}
{"timestamp": "2025-04-12T08:02:35.358728-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 0.0005, "epochs":
10, "batch_size": 32, "weight_decay": 0, "x_task": "single", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,

```

```

"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.10697869211435318, "train_eval_accuracy":
0.9472857142857143, "train_eval_precision": 0.9428081212467829,
"train_eval_recall": 0.9512406231967686, "train_eval_f1": 0.9470056010340371,
"train_eval_runtime": 16.99, "train_eval_samples_per_second": 412.007,
"train_eval_steps_per_second": 12.89, "train_epoch": 10.0, "val_eval_loss":
2.302985429763794, "val_eval_accuracy": 0.592, "val_eval_precision":
0.5764462809917356, "val_eval_recall": 0.578838174273859, "val_eval_f1":
0.577639751552795, "val_eval_runtime": 3.3814, "val_eval_samples_per_second":
295.737, "val_eval_steps_per_second": 9.464, "val_epoch": 10.0,
"test_eval_loss": 2.50677490234375, "test_eval_accuracy": 0.572,
"test_eval_precision": 0.5567226890756303, "test_eval_recall":
0.549792531120332, "test_eval_f1": 0.5532359081419624, "test_eval_runtime":
3.495, "test_eval_samples_per_second": 286.122, "test_eval_steps_per_second":
9.156, "test_epoch": 10.0}}
{"timestamp": "2025-04-12T09:27:28.661745-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 0.0005, "epochs":
25, "batch_size": 16, "weight_decay": 0.05, "x_task": "single", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 5607938,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,

```

```

"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.07316441833972931, "train_eval_accuracy":
0.9521428571428572, "train_eval_precision": 0.9349819394276188,
"train_eval_recall": 0.9708597807270629, "train_eval_f1": 0.9525831564048125,
"train_eval_runtime": 20.7869, "train_eval_samples_per_second": 336.75,
"train_eval_steps_per_second": 21.071, "train_epoch": 25.0, "val_eval_loss":
3.6118979454040527, "val_eval_accuracy": 0.556, "val_eval_precision":
0.5387755102040817, "val_eval_recall": 0.5477178423236515, "val_eval_f1":
0.5432098765432098, "val_eval_runtime": 5.317, "val_eval_samples_per_second":
188.075, "val_eval_steps_per_second": 11.849, "val_epoch": 25.0,
"test_eval_loss": 3.402284860610962, "test_eval_accuracy": 0.555,
"test_eval_precision": 0.5384615384615384, "test_eval_recall":
0.5373443983402489, "test_eval_f1": 0.5379023883696781, "test_eval_runtime":
3.7452, "test_eval_samples_per_second": 267.006, "test_eval_steps_per_second":
16.821, "test_epoch": 25.0}}
{"timestamp": "2025-04-12T09:33:04.239733-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 0.0005, "epochs":
5, "batch_size": 16, "weight_decay": 0.01, "x_task": "single", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.16.attn_norm.weight", "model.layers.16.attn.Wqkv.weight",
"model.layers.16.attn.Wo.weight", "model.layers.16.mlp_norm.weight",
"model.layers.16.mlp.Wi.weight", "model.layers.16.mlp.Wo.weight",
"model.layers.17.attn_norm.weight", "model.layers.17.attn.Wqkv.weight",

```

```

"model.layers.17.attn.Wo.weight", "model.layers.17.mlp_norm.weight",
"model.layers.17.mlp.Wi.weight", "model.layers.17.mlp.Wo.weight",
"model.layers.18.attn_norm.weight", "model.layers.18.attn.Wqkv.weight",
"model.layers.18.attn.Wo.weight", "model.layers.18.mlp_norm.weight",
"model.layers.18.mlp.Wi.weight", "model.layers.18.mlp.Wo.weight",
"model.layers.19.attn_norm.weight", "model.layers.19.attn.Wqkv.weight",
"model.layers.19.attn.Wo.weight", "model.layers.19.mlp_norm.weight",
"model.layers.19.mlp.Wi.weight", "model.layers.19.mlp.Wo.weight",
"model.layers.20.attn_norm.weight", "model.layers.20.attn.Wqkv.weight",
"model.layers.20.attn.Wo.weight", "model.layers.20.mlp_norm.weight",
"model.layers.20.mlp.Wi.weight", "model.layers.20.mlp.Wo.weight",
"model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 30683138,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,

```

```

"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.26596900820732117, "train_eval_accuracy": 0.895,
"train_eval_precision": 0.8731893960098388, "train_eval_recall":
0.9218118869013272, "train_eval_f1": 0.8968421052631579, "train_eval_runtime":
18.8056, "train_eval_samples_per_second": 372.229,
"train_eval_steps_per_second": 23.291, "train_epoch": 5.0, "val_eval_loss":
0.8603283762931824, "val_eval_accuracy": 0.616, "val_eval_precision":
0.5983935742971888, "val_eval_recall": 0.6182572614107884, "val_eval_f1":
0.6081632653061224, "val_eval_runtime": 4.6237, "val_eval_samples_per_second":
216.278, "val_eval_steps_per_second": 13.626, "val_epoch": 5.0,
"test_eval_loss": 0.9077507853507996, "test_eval_accuracy": 0.582,
"test_eval_precision": 0.5677966101694916, "test_eval_recall":
0.5560165975103735, "test_eval_f1": 0.5618448637316562, "test_eval_runtime":
3.8279, "test_eval_samples_per_second": 261.239, "test_eval_steps_per_second":
16.458, "test_epoch": 5.0}}
{"timestamp": "2025-04-12T10:00:57.400798-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 0.0005, "epochs":
10, "batch_size": 4, "weight_decay": 0.0025, "x_task": "single", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.16.attn_norm.weight", "model.layers.16.attn.Wqkv.weight",
"model.layers.16.attn.Wo.weight", "model.layers.16.mlp_norm.weight",
"model.layers.16.mlp.Wi.weight", "model.layers.16.mlp.Wo.weight",
"model.layers.17.attn_norm.weight", "model.layers.17.attn.Wqkv.weight",
"model.layers.17.attn.Wo.weight", "model.layers.17.mlp_norm.weight",
"model.layers.17.mlp.Wi.weight", "model.layers.17.mlp.Wo.weight",
"model.layers.18.attn_norm.weight", "model.layers.18.attn.Wqkv.weight",
"model.layers.18.attn.Wo.weight", "model.layers.18.mlp_norm.weight",
"model.layers.18.mlp.Wi.weight", "model.layers.18.mlp.Wo.weight",
"model.layers.19.attn_norm.weight", "model.layers.19.attn.Wqkv.weight",
"model.layers.19.attn.Wo.weight", "model.layers.19.mlp_norm.weight",
"model.layers.19.mlp.Wi.weight", "model.layers.19.mlp.Wo.weight",
"model.layers.20.attn_norm.weight", "model.layers.20.attn.Wqkv.weight",
"model.layers.20.attn.Wo.weight", "model.layers.20.mlp_norm.weight",
"model.layers.20.mlp.Wi.weight", "model.layers.20.mlp.Wo.weight",
"model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 30683138,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,

```

```

"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.4228428900241852, "train_eval_accuracy":
0.8411428571428572, "train_eval_precision": 0.8187973997833152,
"train_eval_recall": 0.8721869590305829, "train_eval_f1": 0.8446493433920089,
"train_eval_runtime": 55.6925, "train_eval_samples_per_second": 125.69,
"train_eval_steps_per_second": 31.423, "train_epoch": 10.0, "val_eval_loss":
1.1260782480239868, "val_eval_accuracy": 0.602, "val_eval_precision":
0.5789473684210527, "val_eval_recall": 0.6390041493775933, "val_eval_f1":
0.6074950690335306, "val_eval_runtime": 8.984, "val_eval_samples_per_second":
111.309, "val_eval_steps_per_second": 27.827, "val_epoch": 10.0,
"test_eval_loss": 1.2696013450622559, "test_eval_accuracy": 0.575,
"test_eval_precision": 0.5547024952015355, "test_eval_recall":
0.5995850622406639, "test_eval_f1": 0.576271186440678, "test_eval_runtime":
8.9038, "test_eval_samples_per_second": 112.312, "test_eval_steps_per_second":
28.078, "test_epoch": 10.0}}

```



```

{"timestamp": "2025-04-12T10:56:47.131248-07:00", "experiment_meta":
{"model_name": "answerdotai/ModernBERT-base", "learning_rate": 0.075, "epochs":
7, "batch_size": 2, "weight_decay": 0.0025, "x_task": "single", "x_col":
"snc_pos_seq", "y_col": "binary_complexity", "layers_to_unfreeze":
["model.layers.16.attn_norm.weight", "model.layers.16.attn.Wqkv.weight",
"model.layers.16.attn.Wo.weight", "model.layers.16.mlp_norm.weight",
"model.layers.16.mlp.Wi.weight", "model.layers.16.mlp.Wo.weight",
"model.layers.17.attn_norm.weight", "model.layers.17.attn.Wqkv.weight",
"model.layers.17.attn.Wo.weight", "model.layers.17.mlp_norm.weight",
"model.layers.17.mlp.Wi.weight", "model.layers.17.mlp.Wo.weight",
"model.layers.18.attn_norm.weight", "model.layers.18.attn.Wqkv.weight",
"model.layers.18.attn.Wo.weight", "model.layers.18.mlp_norm.weight",
"model.layers.18.mlp.Wi.weight", "model.layers.18.mlp.Wo.weight",
"model.layers.19.attn_norm.weight", "model.layers.19.attn.Wqkv.weight",
"model.layers.19.attn.Wo.weight", "model.layers.19.mlp_norm.weight",
"model.layers.19.mlp.Wi.weight", "model.layers.19.mlp.Wo.weight",
"model.layers.20.attn_norm.weight", "model.layers.20.attn.Wqkv.weight",
"model.layers.20.attn.Wo.weight", "model.layers.20.mlp_norm.weight",
"model.layers.20.mlp.Wi.weight", "model.layers.20.mlp.Wo.weight",
"model.layers.21.attn_norm.weight", "model.layers.21.attn.Wqkv.weight",
"model.layers.21.attn.Wo.weight", "model.layers.21.mlp_norm.weight",
"model.layers.21.mlp.Wi.weight", "model.layers.21.mlp.Wo.weight",
"model.final_norm.weight", "head.dense.weight", "head.norm.weight",
"classifier.weight", "classifier.bias"]}, "model_details":
{"model_total_params": 149606402, "model_trainable_params": 30683138,
"pooler_activation_function": "N/A", "config_attributes": {"return_dict": true,
"output_hidden_states": false, "output_attentions": false, "torchscript": false,
"torch_dtype": "float32", "use_bfloat16": false, "tf_legacy_loss": false,
"pruned_heads": {}, "tie_word_embeddings": true, "chunk_size_feed_forward": 0,
"is_encoder_decoder": false, "is_decoder": false, "cross_attention_hidden_size":
null, "add_cross_attention": false, "tie_encoder_decoder": false, "max_length":
20, "min_length": 0, "do_sample": false, "early_stopping": false, "num_beams":
1, "num_beam_groups": 1, "diversity_penalty": 0.0, "temperature": 1.0, "top_k":
50, "top_p": 1.0, "typical_p": 1.0, "repetition_penalty": 1.0, "length_penalty":
1.0, "no_repeat_ngram_size": 0, "encoder_no_repeat_ngram_size": 0,
"bad_words_ids": null, "num_return_sequences": 1, "output_scores": false,
"return_dict_in_generate": false, "forced_bos_token_id": null,
"forced_eos_token_id": null, "remove_invalid_values": false,
"exponential_decay_length_penalty": null, "suppress_tokens": null,
"begin_suppress_tokens": null, "architectures":
["ModernBertForSequenceClassification"], "finetuning_task": null, "id2label":
{"0": "LABEL_0", "1": "LABEL_1"}, "label2id": {"LABEL_0": 0, "LABEL_1": 1},
"tokenizer_class": null, "prefix": null, "bos_token_id": 50281, "pad_token_id":
50283, "eos_token_id": 50282, "sep_token_id": 50282, "decoder_start_token_id":
null, "task_specific_params": null, "problem_type":
"single_label_classification", "_name_or_path": "answerdotai/ModernBERT-base",
"_commit_hash": "8949b909ec900327062f0ebf497f51aef5e6f0c8",
"_attn_implementation_internal": "sdpa", "_attn_implementation_autoset": false,

```

```
"transformers_version": "4.47.0.dev0", "cls_token_id": 50281,
"gradient_checkpointing": false, "layer_norm_eps": 1e-05, "model_type":
"modernbert", "position_embedding_type": "absolute", "vocab_size": 50368,
"max_position_embeddings": 8192, "hidden_size": 768, "intermediate_size": 1152,
"num_hidden_layers": 22, "num_attention_heads": 12, "initializer_range": 0.02,
"initializer_cutoff_factor": 2.0, "norm_eps": 1e-05, "norm_bias": false,
"global_rope_theta": 160000.0, "attention_bias": false, "attention_dropout":
0.0, "hidden_activation": "gelu", "global_attn_every_n_layers": 3,
"local_attention": 128, "local_rope_theta": 10000.0, "embedding_dropout": 0.0,
"mlp_bias": false, "mlp_dropout": 0.0, "decoder_bias": true,
"classifier_pooling": "mean", "classifier_dropout": 0.0, "classifier_bias":
false, "classifier_activation": "gelu", "deterministic_flash_attn": false,
"sparse_prediction": false, "sparse_pred_ignore_index": -100,
"reference_compile": true, "repad_logits_with_grad": false}}, "run_metrics":
{"train_eval_loss": 0.6931136846542358, "train_eval_accuracy":
0.5048571428571429, "train_eval_precision": 0.0, "train_eval_recall": 0.0,
"train_eval_f1": 0.0, "train_eval_runtime": 112.3633,
"train_eval_samples_per_second": 62.298, "train_eval_steps_per_second": 31.149,
"train_epoch": 7.0, "val_eval_loss": 0.6927208304405212, "val_eval_accuracy":
0.518, "val_eval_precision": 0.0, "val_eval_recall": 0.0, "val_eval_f1": 0.0,
"val_eval_runtime": 16.7576, "val_eval_samples_per_second": 59.674,
"val_eval_steps_per_second": 29.837, "val_epoch": 7.0, "test_eval_loss":
0.692720890045166, "test_eval_accuracy": 0.518, "test_eval_precision": 0.0,
"test_eval_recall": 0.0, "test_eval_f1": 0.0, "test_eval_runtime": 17.0029,
"test_eval_samples_per_second": 58.813, "test_eval_steps_per_second": 29.407,
"test_epoch": 7.0}}
```

```
[14]: print(log_df.columns)
```

```
Index(['timestamp', 'experiment_meta.model_name',
      'experiment_meta.learning_rate', 'experiment_meta.epochs',
      'experiment_meta.batch_size', 'experiment_meta.weight_decay',
      'experiment_meta.x_task', 'experiment_meta.x_col',
      'experiment_meta.y_col', 'experiment_meta.layers_to_unfreeze',
      ...
      'model_details.config_attributes.summary_type',
      'model_details.config_attributes.summary_use_proj',
      'model_details.config_attributes.summary_activation',
      'model_details.config_attributes.summary_last_dropout',
      'model_details.config_attributes.start_n_top',
      'model_details.config_attributes.end_n_top',
      'model_details.config_attributes.use_mems_eval',
      'model_details.config_attributes.use_mems_train',
      'model_details.config_attributes.task_specific_params.text-
generation.do_sample',
      'model_details.config_attributes.task_specific_params.text-
generation.max_length'],
```

```
dtype='object', length=180)
```

```
[15]: # prompt: THIS OUTPUT IS TRUNCATING THE COLUMN, WRITE THE COLUMNS TO A LIST AND  
      ↪ THEN PRINT THAT LIST
```

```
column_list = log_df.columns.tolist()  
column_list
```

```
[15]: ['timestamp',  
      'experiment_meta.model_name',  
      'experiment_meta.learning_rate',  
      'experiment_meta.epochs',  
      'experiment_meta.batch_size',  
      'experiment_meta.weight_decay',  
      'experiment_meta.x_task',  
      'experiment_meta.x_col',  
      'experiment_meta.y_col',  
      'experiment_meta.layers_to_unfreeze',  
      'model_details.model_total_params',  
      'model_details.model_trainable_params',  
      'model_details.pooler_activation_function',  
      'model_details.config_attributes.return_dict',  
      'model_details.config_attributes.output_hidden_states',  
      'model_details.config_attributes.output attentions',  
      'model_details.config_attributes.torchscript',  
      'model_details.config_attributes.torch_dtype',  
      'model_details.config_attributes.use_bfloat16',  
      'model_details.config_attributes.tf_legacy_loss',  
      'model_details.config_attributes.tie_word_embeddings',  
      'model_details.config_attributes.chunk_size_feed_forward',  
      'model_details.config_attributes.is_encoder_decoder',  
      'model_details.config_attributes.is_decoder',  
      'model_details.config_attributes.cross_attention_hidden_size',  
      'model_details.config_attributes.add_cross_attention',  
      'model_details.config_attributes.tie_encoder_decoder',  
      'model_details.config_attributes.max_length',  
      'model_details.config_attributes.min_length',  
      'model_details.config_attributes.do_sample',  
      'model_details.config_attributes.early_stopping',  
      'model_details.config_attributes.num_beams',  
      'model_details.config_attributes.num_beam_groups',  
      'model_details.config_attributes.diversity_penalty',  
      'model_details.config_attributes.temperature',  
      'model_details.config_attributes.top_k',  
      'model_details.config_attributes.top_p',  
      'model_details.config_attributes.typical_p',  
      'model_details.config_attributes.repetition_penalty',
```

'model_details.config_attributes.length_penalty',
 'model_details.config_attributes.no_repeat_ngram_size',
 'model_details.config_attributes.encoder_no_repeat_ngram_size',
 'model_details.config_attributes.bad_words_ids',
 'model_details.config_attributes.num_return_sequences',
 'model_details.config_attributes.output_scores',
 'model_details.config_attributes.return_dict_in_generate',
 'model_details.config_attributes.forced_bos_token_id',
 'model_details.config_attributes.forced_eos_token_id',
 'model_details.config_attributes.remove_invalid_values',
 'model_details.config_attributes.exponential_decay_length_penalty',
 'model_details.config_attributes.suppress_tokens',
 'model_details.config_attributes.begin_suppress_tokens',
 'model_details.config_attributes.architectures',
 'model_details.config_attributes.finetuning_task',
 'model_details.config_attributes.id2label.0',
 'model_details.config_attributes.id2label.1',
 'model_details.config_attributes.label2id.LABEL_0',
 'model_details.config_attributes.label2id.LABEL_1',
 'model_details.config_attributes.tokenizer_class',
 'model_details.config_attributes.prefix',
 'model_details.config_attributes.bos_token_id',
 'model_details.config_attributes.pad_token_id',
 'model_details.config_attributes.eos_token_id',
 'model_details.config_attributes.sep_token_id',
 'model_details.config_attributes.decoder_start_token_id',
 'model_details.config_attributes.task_specific_params',
 'model_details.config_attributes.problem_type',
 'model_details.config_attributes._name_or_path',
 'model_details.config_attributes._commit_hash',
 'model_details.config_attributes._attn_implementation_internal',
 'model_details.config_attributes._attn_implementation_autoset',
 'model_details.config_attributes.transformers_version',
 'model_details.config_attributes.gradient_checkpointing',
 'model_details.config_attributes.model_type',
 'model_details.config_attributes.vocab_size',
 'model_details.config_attributes.hidden_size',
 'model_details.config_attributes.num_hidden_layers',
 'model_details.config_attributes.num_attention_heads',
 'model_details.config_attributes.hidden_act',
 'model_details.config_attributes.intermediate_size',
 'model_details.config_attributes.hidden_dropout_prob',
 'model_details.config_attributes.attention_probs_dropout_prob',
 'model_details.config_attributes.max_position_embeddings',
 'model_details.config_attributes.type_vocab_size',
 'model_details.config_attributes.initializer_range',
 'model_details.config_attributes.layer_norm_eps',

```

'model_details.config_attributes.position_embedding_type',
'model_details.config_attributes.use_cache',
'model_details.config_attributes.classifier_dropout',
'run_metrics.train_eval_loss',
'run_metrics.train_eval_accuracy',
'run_metrics.train_eval_precision',
'run_metrics.train_eval_recall',
'run_metrics.train_eval_f1',
'run_metrics.train_eval_runtime',
'run_metrics.train_eval_samples_per_second',
'run_metrics.train_eval_steps_per_second',
'run_metrics.train_epoch',
'run_metrics.val_eval_loss',
'run_metrics.val_eval_accuracy',
'run_metrics.val_eval_precision',
'run_metrics.val_eval_recall',
'run_metrics.val_eval_f1',
'run_metrics.val_eval_runtime',
'run_metrics.val_eval_samples_per_second',
'run_metrics.val_eval_steps_per_second',
'run_metrics.val_epoch',
'run_metrics.test_eval_loss',
'run_metrics.test_eval_accuracy',
'run_metrics.test_eval_precision',
'run_metrics.test_eval_recall',
'run_metrics.test_eval_f1',
'run_metrics.test_eval_runtime',
'run_metrics.test_eval_samples_per_second',
'run_metrics.test_eval_steps_per_second',
'run_metrics.test_epoch',
'model_details.config_attributes.directionality',
'model_details.config_attributes.pooler_fc_size',
'model_details.config_attributes.pooler_num_attention_heads',
'model_details.config_attributes.pooler_num_fc_layers',
'model_details.config_attributes.pooler_size_per_head',
'model_details.config_attributes.pooler_type',
'model_details.config_attributes.cls_token_id',
'model_details.config_attributes.initializer_cutoff_factor',
'model_details.config_attributes.norm_eps',
'model_details.config_attributes.norm_bias',
'model_details.config_attributes.global_rope_theta',
'model_details.config_attributes.attention_bias',
'model_details.config_attributes.attention_dropout',
'model_details.config_attributes.hidden_activation',
'model_details.config_attributes.global_attn_every_n_layers',
'model_details.config_attributes.local_attention',
'model_details.config_attributes.local_rope_theta',

```

```

'model_details.config_attributes.embedding_dropout',
'model_details.config_attributes.mlp_bias',
'model_details.config_attributes.mlp_dropout',
'model_details.config_attributes.decoder_bias',
'model_details.config_attributes.classifier_pooling',
'model_details.config_attributes.classifier_bias',
'model_details.config_attributes.classifier_activation',
'model_details.config_attributes.deterministic_flash_attn',
'model_details.config_attributes.sparse_prediction',
'model_details.config_attributes.sparse_pred_ignore_index',
'model_details.config_attributes.reference_compile',
'model_details.config_attributes.repad_logits_with_grad',
'model_details.config_attributes.position_buckets',
'model_details.config_attributes.norm_rel_ebd',
'model_details.config_attributes.share_att_key',
'model_details.config_attributes.relative_attention',
'model_details.config_attributes.max_relative_positions',
'model_details.config_attributes.position_biased_input',
'model_details.config_attributes.pos_att_type',
'model_details.config_attributes.pooler_hidden_size',
'model_details.config_attributes.pooler_dropout',
'model_details.config_attributes.pooler_hidden_act',
'model_details.config_attributes.legacy',
'model_details.config_attributes.d_model',
'model_details.config_attributes.n_layer',
'model_details.config_attributes.n_head',
'model_details.config_attributes.d_head',
'model_details.config_attributes.ff_activation',
'model_details.config_attributes.d_inner',
'model_details.config_attributes.untie_r',
'model_details.config_attributes.attn_type',
'model_details.config_attributes.dropout',
'model_details.config_attributes.mem_len',
'model_details.config_attributes.reuse_len',
'model_details.config_attributes.bi_data',
'model_details.config_attributes.clamp_len',
'model_details.config_attributes.same_length',
'model_details.config_attributes.summary_type',
'model_details.config_attributes.summary_use_proj',
'model_details.config_attributes.summary_activation',
'model_details.config_attributes.summary_last_dropout',
'model_details.config_attributes.start_n_top',
'model_details.config_attributes.end_n_top',
'model_details.config_attributes.use_mems_eval',
'model_details.config_attributes.use_mems_train',
'model_details.config_attributes.task_specific_params.text-
generation.do_sample',

```

```
'model_details.config_attributes.task_specific_params.text-  
generation.max_length']
```

```
[16]: def display_experiment_metrics(log_df):  
    base_columns = [  
        'timestamp',  
        'experiment_meta.model_name',  
        'experiment_meta.x_task',  
        'experiment_meta.x_col',  
        # 'model_lineage.type',  
        # 'model_lineage.path',  
        # 'model_lineage.timestamp',  
        # 'model_details.config_attributes._name_or_path',  
        # 'model_details.config_attributes.classifier_dropout',  
        'model_details.config_attributes.attention_probs_dropout_prob',  
        'model_details.config_attributes.hidden_dropout_prob',  
        'model_details.config_attributes.num_hidden_layers',  
        'model_details.model_total_params',  
        'model_details.model_trainable_params',  
        'experiment_meta.model_name',  
        'experiment_meta.learning_rate',  
        'experiment_meta.epochs',  
        'experiment_meta.batch_size',  
        'experiment_meta.weight_decay',  
        'experiment_meta.x_task',  
        'experiment_meta.x_col',  
        'experiment_meta.y_col',  
        'experiment_meta.layers_to_unfreeze',  
    ]  
  
    metrics = [  
        'run_metrics.train_eval_loss',  
        'run_metrics.train_eval_accuracy',  
        'run_metrics.train_eval_precision',  
        'run_metrics.train_eval_recall',  
        'run_metrics.train_eval_f1',  
        'run_metrics.train_eval_runtime',  
        'run_metrics.train_eval_samples_per_second',  
        'run_metrics.train_eval_steps_per_second',  
        'run_metrics.train_epoch',  
        'run_metrics.val_eval_loss',  
        'run_metrics.val_eval_accuracy',  
        'run_metrics.val_eval_precision',  
        'run_metrics.val_eval_recall',  
        'run_metrics.val_eval_f1',  
        'run_metrics.val_eval_runtime',
```

```

        'run_metrics.val_eval_samples_per_second',
        'run_metrics.val_eval_steps_per_second',
        # 'run_metrics.val_epoch',
        'run_metrics.test_eval_loss',
        'run_metrics.test_eval_accuracy',
        'run_metrics.test_eval_precision',
        'run_metrics.test_eval_recall',
        'run_metrics.test_eval_f1',
        'run_metrics.test_eval_runtime',
        'run_metrics.test_eval_samples_per_second',
        'run_metrics.test_eval_steps_per_second',
        # 'run_metrics.test_epoch'
    ]

    desired_columns = base_columns + metrics

    display_df = log_df.reindex(columns=desired_columns)

    return display_df

result_df = display_experiment_metrics(log_df)
display(result_df)

```

| | timestamp | experiment_meta.model_name \ |
|-----|----------------------------------|------------------------------|
| 0 | 2025-04-09T10:57:21.454889-07:00 | bert-base-cased |
| 1 | 2025-04-09T10:58:27.004579-07:00 | bert-base-cased |
| 2 | 2025-04-09T10:59:19.669997-07:00 | bert-base-cased |
| 3 | 2025-04-09T11:00:14.558023-07:00 | bert-base-cased |
| 4 | 2025-04-09T11:01:07.013948-07:00 | bert-base-cased |
| .. | ... | ... |
| 243 | 2025-04-12T08:02:35.358728-07:00 | answerdotai/ModernBERT-base |
| 244 | 2025-04-12T09:27:28.661745-07:00 | answerdotai/ModernBERT-base |
| 245 | 2025-04-12T09:33:04.239733-07:00 | answerdotai/ModernBERT-base |
| 246 | 2025-04-12T10:00:57.400798-07:00 | answerdotai/ModernBERT-base |
| 247 | 2025-04-12T10:56:47.131248-07:00 | answerdotai/ModernBERT-base |

| | experiment_meta.x_task | experiment_meta.x_col \ |
|-----|------------------------|-------------------------|
| 0 | single | pos_sequence |
| 1 | multi | pos_sequence |
| 2 | single | morph_sequence |
| 3 | multi | morph_sequence |
| 4 | single | sentence |
| .. | ... | ... |
| 243 | single | snc_pos_seq |
| 244 | single | snc_pos_seq |
| 245 | single | snc_pos_seq |
| 246 | single | snc_pos_seq |
| 247 | single | snc_pos_seq |

| | model_details.config_attributes.attention_probs_dropout_prob \ |
|-----|--|
| 0 | 0.1 |
| 1 | 0.1 |
| 2 | 0.1 |
| 3 | 0.1 |
| 4 | 0.1 |
| .. | ... |
| 243 | NaN |
| 244 | NaN |
| 245 | NaN |
| 246 | NaN |
| 247 | NaN |

| | model_details.config_attributes.hidden_dropout_prob \ |
|-----|---|
| 0 | 0.1 |
| 1 | 0.1 |
| 2 | 0.1 |
| 3 | 0.1 |
| 4 | 0.1 |
| .. | ... |
| 243 | NaN |
| 244 | NaN |
| 245 | NaN |
| 246 | NaN |
| 247 | NaN |

| | model_details.config_attributes.num_hidden_layers \ |
|-----|---|
| 0 | 12.0 |
| 1 | 12.0 |
| 2 | 12.0 |
| 3 | 12.0 |
| 4 | 12.0 |
| .. | ... |
| 243 | 22.0 |
| 244 | 22.0 |
| 245 | 22.0 |
| 246 | 22.0 |
| 247 | 22.0 |

| | model_details.model_total_params | model_details.model_trainable_params \ |
|-----|----------------------------------|--|
| 0 | 108311810 | 14767874 |
| 1 | 108311810 | 14767874 |
| 2 | 108311810 | 14767874 |
| 3 | 108311810 | 14767874 |
| 4 | 108311810 | 14767874 |
| .. | ... | ... |
| 243 | 149606402 | 5607938 |

| | | |
|-----|-----------|----------|
| 244 | 149606402 | 5607938 |
| 245 | 149606402 | 30683138 |
| 246 | 149606402 | 30683138 |
| 247 | 149606402 | 30683138 |

| | | | |
|-----|-----------------------------|-----|---|
| | experiment_meta.model_name | ... | \ |
| 0 | bert-base-cased | ... | |
| 1 | bert-base-cased | ... | |
| 2 | bert-base-cased | ... | |
| 3 | bert-base-cased | ... | |
| 4 | bert-base-cased | ... | |
| .. | ... | ... | |
| 243 | answerdotai/ModernBERT-base | ... | |
| 244 | answerdotai/ModernBERT-base | ... | |
| 245 | answerdotai/ModernBERT-base | ... | |
| 246 | answerdotai/ModernBERT-base | ... | |
| 247 | answerdotai/ModernBERT-base | ... | |

| | | |
|-----|---|---|
| | run_metrics.val_eval_samples_per_second | \ |
| 0 | 195.437 | |
| 1 | 132.559 | |
| 2 | 192.625 | |
| 3 | 176.040 | |
| 4 | 130.929 | |
| .. | ... | |
| 243 | 295.737 | |
| 244 | 188.075 | |
| 245 | 216.278 | |
| 246 | 111.309 | |
| 247 | 59.674 | |

| | | | |
|-----|---------------------------------------|----------------------------|---|
| | run_metrics.val_eval_steps_per_second | run_metrics.test_eval_loss | \ |
| 0 | 1.857 | 0.690998 | |
| 1 | 1.259 | 0.686096 | |
| 2 | 1.830 | 0.714042 | |
| 3 | 1.673 | 0.714042 | |
| 4 | 1.244 | 0.714042 | |
| .. | ... | ... | |
| 243 | 9.464 | 2.506775 | |
| 244 | 11.849 | 3.402285 | |
| 245 | 13.626 | 0.907751 | |
| 246 | 27.827 | 1.269601 | |
| 247 | 29.837 | 0.692721 | |

| | | | |
|---|--------------------------------|---------------------------------|---|
| | run_metrics.test_eval_accuracy | run_metrics.test_eval_precision | \ |
| 0 | 0.520174 | 0.500726 | |
| 1 | 0.556161 | 0.555921 | |
| 2 | 0.480916 | 0.480916 | |

| | | |
|-----|----------|----------|
| 3 | 0.480916 | 0.480916 |
| 4 | 0.480916 | 0.480916 |
| .. | ... | ... |
| 243 | 0.572000 | 0.556723 |
| 244 | 0.555000 | 0.538462 |
| 245 | 0.582000 | 0.567797 |
| 246 | 0.575000 | 0.554702 |
| 247 | 0.518000 | 0.000000 |

| | run_metrics.test_eval_recall | run_metrics.test_eval_f1 \ |
|-----|------------------------------|----------------------------|
| 0 | 0.782313 | 0.610619 |
| 1 | 0.383220 | 0.453691 |
| 2 | 1.000000 | 0.649485 |
| 3 | 1.000000 | 0.649485 |
| 4 | 1.000000 | 0.649485 |
| .. | ... | ... |
| 243 | 0.549793 | 0.553236 |
| 244 | 0.537344 | 0.537902 |
| 245 | 0.556017 | 0.561845 |
| 246 | 0.599585 | 0.576271 |
| 247 | 0.000000 | 0.000000 |

| | run_metrics.test_eval_runtime | run_metrics.test_eval_samples_per_second \ |
|-----|-------------------------------|--|
| 0 | 2.6416 | 347.136 |
| 1 | 4.7886 | 191.498 |
| 2 | 2.8863 | 317.704 |
| 3 | 3.0385 | 301.796 |
| 4 | 3.9224 | 233.785 |
| .. | ... | ... |
| 243 | 3.4950 | 286.122 |
| 244 | 3.7452 | 267.006 |
| 245 | 3.8279 | 261.239 |
| 246 | 8.9038 | 112.312 |
| 247 | 17.0029 | 58.813 |

| | run_metrics.test_eval_steps_per_second |
|-----|--|
| 0 | 3.028 |
| 1 | 1.671 |
| 2 | 2.772 |
| 3 | 2.633 |
| 4 | 2.040 |
| .. | ... |
| 243 | 9.156 |
| 244 | 16.821 |
| 245 | 16.458 |
| 246 | 28.078 |
| 247 | 29.407 |

[248 rows x 43 columns]

```
[17]: # export as xlsx excel file

result_df = display_experiment_metrics(log_df)
display(result_df)

# Create the results directory if it doesn't exist
os.makedirs(dir_results, exist_ok=True)

# Generate timestamp for filename

pacific_time = datetime.now(zoneinfo.ZoneInfo("America/Los_Angeles"))
timestamp = pacific_time.isoformat()
filename = f"experiment_results_{timestamp}.xlsx"
filepath = os.path.join(dir_results, filename)

# Export the DataFrame to an Excel file
result_df.to_excel(filepath, index=False)

print(f"Results exported to: {filepath}")
```

| | timestamp | experiment_meta.model_name \ |
|-----|----------------------------------|------------------------------|
| 0 | 2025-04-09T10:57:21.454889-07:00 | bert-base-cased |
| 1 | 2025-04-09T10:58:27.004579-07:00 | bert-base-cased |
| 2 | 2025-04-09T10:59:19.669997-07:00 | bert-base-cased |
| 3 | 2025-04-09T11:00:14.558023-07:00 | bert-base-cased |
| 4 | 2025-04-09T11:01:07.013948-07:00 | bert-base-cased |
| .. | ... | ... |
| 243 | 2025-04-12T08:02:35.358728-07:00 | answerdotai/ModernBERT-base |
| 244 | 2025-04-12T09:27:28.661745-07:00 | answerdotai/ModernBERT-base |
| 245 | 2025-04-12T09:33:04.239733-07:00 | answerdotai/ModernBERT-base |
| 246 | 2025-04-12T10:00:57.400798-07:00 | answerdotai/ModernBERT-base |
| 247 | 2025-04-12T10:56:47.131248-07:00 | answerdotai/ModernBERT-base |

| | experiment_meta.x_task | experiment_meta.x_col \ |
|-----|------------------------|-------------------------|
| 0 | single | pos_sequence |
| 1 | multi | pos_sequence |
| 2 | single | morph_sequence |
| 3 | multi | morph_sequence |
| 4 | single | sentence |
| .. | ... | ... |
| 243 | single | snc_pos_seq |
| 244 | single | snc_pos_seq |
| 245 | single | snc_pos_seq |
| 246 | single | snc_pos_seq |
| 247 | single | snc_pos_seq |

| | model_details.config_attributes.attention_probs_dropout_prob \ | |
|-----|--|--|
| 0 | 0.1 | |
| 1 | 0.1 | |
| 2 | 0.1 | |
| 3 | 0.1 | |
| 4 | 0.1 | |
| .. | ... | |
| 243 | NaN | |
| 244 | NaN | |
| 245 | NaN | |
| 246 | NaN | |
| 247 | NaN | |

| | model_details.config_attributes.hidden_dropout_prob \ | |
|-----|---|--|
| 0 | 0.1 | |
| 1 | 0.1 | |
| 2 | 0.1 | |
| 3 | 0.1 | |
| 4 | 0.1 | |
| .. | ... | |
| 243 | NaN | |
| 244 | NaN | |
| 245 | NaN | |
| 246 | NaN | |
| 247 | NaN | |

| | model_details.config_attributes.num_hidden_layers \ | |
|-----|---|--|
| 0 | 12.0 | |
| 1 | 12.0 | |
| 2 | 12.0 | |
| 3 | 12.0 | |
| 4 | 12.0 | |
| .. | ... | |
| 243 | 22.0 | |
| 244 | 22.0 | |
| 245 | 22.0 | |
| 246 | 22.0 | |
| 247 | 22.0 | |

| | model_details.model_total_params | model_details.model_trainable_params \ |
|-----|----------------------------------|--|
| 0 | 108311810 | 14767874 |
| 1 | 108311810 | 14767874 |
| 2 | 108311810 | 14767874 |
| 3 | 108311810 | 14767874 |
| 4 | 108311810 | 14767874 |
| .. | ... | ... |
| 243 | 149606402 | 5607938 |
| 244 | 149606402 | 5607938 |

| | | |
|-----|-----------|----------|
| 245 | 149606402 | 30683138 |
| 246 | 149606402 | 30683138 |
| 247 | 149606402 | 30683138 |

| | | | |
|-----|-----------------------------|-----|---|
| | experiment_meta.model_name | ... | \ |
| 0 | bert-base-cased | ... | |
| 1 | bert-base-cased | ... | |
| 2 | bert-base-cased | ... | |
| 3 | bert-base-cased | ... | |
| 4 | bert-base-cased | ... | |
| .. | ... | ... | |
| 243 | answerdotai/ModernBERT-base | ... | |
| 244 | answerdotai/ModernBERT-base | ... | |
| 245 | answerdotai/ModernBERT-base | ... | |
| 246 | answerdotai/ModernBERT-base | ... | |
| 247 | answerdotai/ModernBERT-base | ... | |

| | | |
|-----|---|---|
| | run_metrics.val_eval_samples_per_second | \ |
| 0 | 195.437 | |
| 1 | 132.559 | |
| 2 | 192.625 | |
| 3 | 176.040 | |
| 4 | 130.929 | |
| .. | ... | |
| 243 | 295.737 | |
| 244 | 188.075 | |
| 245 | 216.278 | |
| 246 | 111.309 | |
| 247 | 59.674 | |

| | | | |
|-----|---------------------------------------|----------------------------|---|
| | run_metrics.val_eval_steps_per_second | run_metrics.test_eval_loss | \ |
| 0 | 1.857 | 0.690998 | |
| 1 | 1.259 | 0.686096 | |
| 2 | 1.830 | 0.714042 | |
| 3 | 1.673 | 0.714042 | |
| 4 | 1.244 | 0.714042 | |
| .. | ... | ... | |
| 243 | 9.464 | 2.506775 | |
| 244 | 11.849 | 3.402285 | |
| 245 | 13.626 | 0.907751 | |
| 246 | 27.827 | 1.269601 | |
| 247 | 29.837 | 0.692721 | |

| | | | |
|---|--------------------------------|---------------------------------|---|
| | run_metrics.test_eval_accuracy | run_metrics.test_eval_precision | \ |
| 0 | 0.520174 | 0.500726 | |
| 1 | 0.556161 | 0.555921 | |
| 2 | 0.480916 | 0.480916 | |
| 3 | 0.480916 | 0.480916 | |

| | | |
|-----|----------|----------|
| 4 | 0.480916 | 0.480916 |
| .. | ... | ... |
| 243 | 0.572000 | 0.556723 |
| 244 | 0.555000 | 0.538462 |
| 245 | 0.582000 | 0.567797 |
| 246 | 0.575000 | 0.554702 |
| 247 | 0.518000 | 0.000000 |

| | run_metrics.test_eval_recall | run_metrics.test_eval_f1 | \ |
|-----|------------------------------|--------------------------|---|
| 0 | 0.782313 | 0.610619 | |
| 1 | 0.383220 | 0.453691 | |
| 2 | 1.000000 | 0.649485 | |
| 3 | 1.000000 | 0.649485 | |
| 4 | 1.000000 | 0.649485 | |
| .. | ... | ... | |
| 243 | 0.549793 | 0.553236 | |
| 244 | 0.537344 | 0.537902 | |
| 245 | 0.556017 | 0.561845 | |
| 246 | 0.599585 | 0.576271 | |
| 247 | 0.000000 | 0.000000 | |

| | run_metrics.test_eval_runtime | run_metrics.test_eval_samples_per_second | \ |
|-----|-------------------------------|--|---|
| 0 | 2.6416 | 347.136 | |
| 1 | 4.7886 | 191.498 | |
| 2 | 2.8863 | 317.704 | |
| 3 | 3.0385 | 301.796 | |
| 4 | 3.9224 | 233.785 | |
| .. | ... | ... | |
| 243 | 3.4950 | 286.122 | |
| 244 | 3.7452 | 267.006 | |
| 245 | 3.8279 | 261.239 | |
| 246 | 8.9038 | 112.312 | |
| 247 | 17.0029 | 58.813 | |

| | run_metrics.test_eval_steps_per_second |
|-----|--|
| 0 | 3.028 |
| 1 | 1.671 |
| 2 | 2.772 |
| 3 | 2.633 |
| 4 | 2.040 |
| .. | ... |
| 243 | 9.156 |
| 244 | 16.821 |
| 245 | 16.458 |
| 246 | 28.078 |
| 247 | 29.407 |

[248 rows x 43 columns]

Results exported to: /content/drive/MyDrive/266-final/results/experiment_results_2025-04-12T15:27:48.297397-07:00.xlsx

```
[18]: result_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 248 entries, 0 to 247
Data columns (total 43 columns):
 #   Column                                                                                               Non-Null
Count  Dtype
---  -
-----
 0   timestamp                                                    248 non-null
object
 1   experiment_meta.model_name                                    248 non-null
object
 2   experiment_meta.x_task                                       248 non-null
object
 3   experiment_meta.x_col                                        248 non-null
object
 4   model_details.config_attributes.attention_probs_dropout_prob 175 non-null
float64
 5   model_details.config_attributes.hidden_dropout_prob          175 non-null
float64
 6   model_details.config_attributes.num_hidden_layers            223 non-null
float64
 7   model_details.model_total_params                             248 non-null
int64
 8   model_details.model_trainable_params                         248 non-null
int64
 9   experiment_meta.model_name                                    248 non-null
object
10   experiment_meta.learning_rate                                248 non-null
float64
11   experiment_meta.epochs                                       248 non-null
int64
12   experiment_meta.batch_size                                    248 non-null
int64
13   experiment_meta.weight_decay                                  248 non-null
float64
14   experiment_meta.x_task                                       248 non-null
object
15   experiment_meta.x_col                                        248 non-null
object
16   experiment_meta.y_col                                        248 non-null
object
17   experiment_meta.layers_to_unfreeze                           248 non-null
object
```


| | | |
|----|---|--------------|
| 18 | run_metrics.train_eval_loss | 248 non-null |
| | float64 | |
| 19 | run_metrics.train_eval_accuracy | 248 non-null |
| | float64 | |
| 20 | run_metrics.train_eval_precision | 248 non-null |
| | float64 | |
| 21 | run_metrics.train_eval_recall | 248 non-null |
| | float64 | |
| 22 | run_metrics.train_eval_f1 | 248 non-null |
| | float64 | |
| 23 | run_metrics.train_eval_runtime | 248 non-null |
| | float64 | |
| 24 | run_metrics.train_eval_samples_per_second | 248 non-null |
| | float64 | |
| 25 | run_metrics.train_eval_steps_per_second | 248 non-null |
| | float64 | |
| 26 | run_metrics.train_epoch | 248 non-null |
| | float64 | |
| 27 | run_metrics.val_eval_loss | 248 non-null |
| | float64 | |
| 28 | run_metrics.val_eval_accuracy | 248 non-null |
| | float64 | |
| 29 | run_metrics.val_eval_precision | 248 non-null |
| | float64 | |
| 30 | run_metrics.val_eval_recall | 248 non-null |
| | float64 | |
| 31 | run_metrics.val_eval_f1 | 248 non-null |
| | float64 | |
| 32 | run_metrics.val_eval_runtime | 248 non-null |
| | float64 | |
| 33 | run_metrics.val_eval_samples_per_second | 248 non-null |
| | float64 | |
| 34 | run_metrics.val_eval_steps_per_second | 248 non-null |
| | float64 | |
| 35 | run_metrics.test_eval_loss | 248 non-null |
| | float64 | |
| 36 | run_metrics.test_eval_accuracy | 248 non-null |
| | float64 | |
| 37 | run_metrics.test_eval_precision | 248 non-null |
| | float64 | |
| 38 | run_metrics.test_eval_recall | 248 non-null |
| | float64 | |
| 39 | run_metrics.test_eval_f1 | 248 non-null |
| | float64 | |
| 40 | run_metrics.test_eval_runtime | 248 non-null |
| | float64 | |
| 41 | run_metrics.test_eval_samples_per_second | 248 non-null |
| | float64 | |

```

42 run_metrics.test_eval_steps_per_second          248 non-null
float64
dtypes: float64(30), int64(4), object(9)
memory usage: 83.4+ KB

```

```

[19]: # Select only float columns and 'train_eval_f1'
numerical_features = result_df.select_dtypes(include=['number'])

# if test_eval_f1 is not numerical for some reason convert to it.
if not pd.api.types.is_numeric_dtype(numerical_features['run_metrics.
↳train_eval_f1']):
    numerical_features['run_metrics.train_eval_f1'] = pd.
↳to_numeric(numerical_features['run_metrics.train_eval_f1'],errors='coerce')

# Calculate correlations with 'train_eval_f1'
correlations = numerical_features.corr()['run_metrics.train_eval_f1']

# Print or display the correlations
print(correlations)
plt.figure(figsize=(25, 25))
sns.heatmap(numerical_features.corr(), annot=True, cmap='coolwarm')
plt.show()

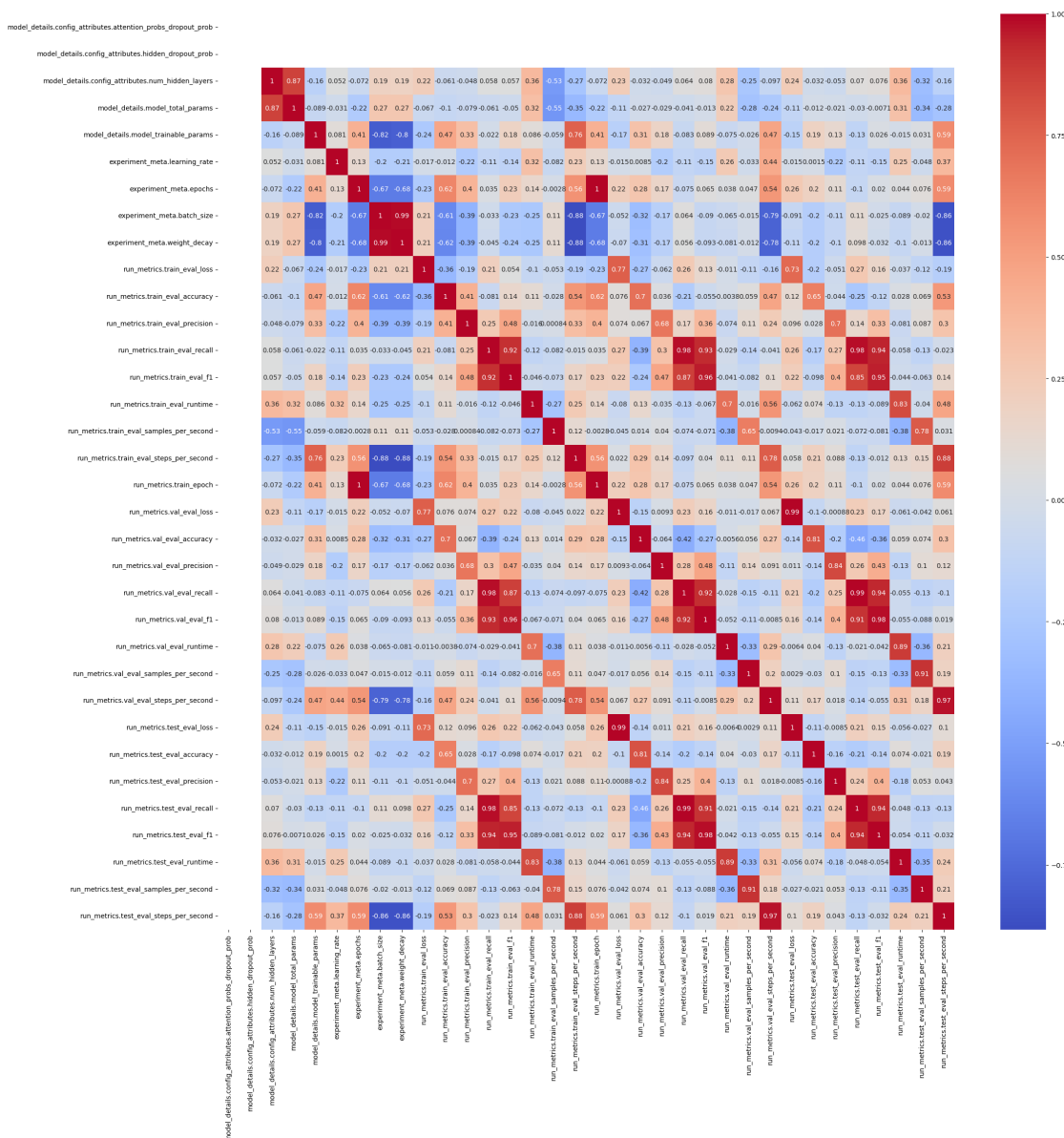
```

```

model_details.config_attributes.attention_probs_dropout_prob    NaN
model_details.config_attributes.hidden_dropout_prob             NaN
model_details.config_attributes.num_hidden_layers               0.056632
model_details.model_total_params                               -0.049522
model_details.model_trainable_params                           0.178747
experiment_meta.learning_rate                                   -0.141809
experiment_meta.epochs                                           0.227186
experiment_meta.batch_size                                       -0.231683
experiment_meta.weight_decay                                     -0.240470
run_metrics.train_eval_loss                                      0.054129
run_metrics.train_eval_accuracy                                 0.142260
run_metrics.train_eval_precision                                0.477430
run_metrics.train_eval_recall                                   0.919058
run_metrics.train_eval_f1                                       1.000000
run_metrics.train_eval_runtime                                   -0.045771
run_metrics.train_eval_samples_per_second                       -0.072789
run_metrics.train_eval_steps_per_second                         0.166722
run_metrics.train_epoch                                          0.227186
run_metrics.val_eval_loss                                        0.216838
run_metrics.val_eval_accuracy                                   -0.239961
run_metrics.val_eval_precision                                  0.467345
run_metrics.val_eval_recall                                     0.871215
run_metrics.val_eval_f1                                         0.962764
run_metrics.val_eval_runtime                                    -0.041091
run_metrics.val_eval_samples_per_second                        -0.082039

```

run_metrics.val_eval_steps_per_second 0.102796
run_metrics.test_eval_loss 0.222344
run_metrics.test_eval_accuracy -0.098481
run_metrics.test_eval_precision 0.404658
run_metrics.test_eval_recall 0.853184
run_metrics.test_eval_f1 0.949814
run_metrics.test_eval_runtime -0.044212
run_metrics.test_eval_samples_per_second -0.063108
run_metrics.test_eval_steps_per_second 0.143413
Name: run_metrics.train_eval_f1, dtype: float64



```
[20]: # Select only float columns and 'val_eval_f1'
numerical_features = result_df.select_dtypes(include=['number'])

# if test_eval_f1 is not numerical for some reason convert to it.
if not pd.api.types.is_numeric_dtype(numerical_features['run_metrics.
    ↪val_eval_f1']):
    numerical_features['run_metrics.val_eval_f1'] = pd.
    ↪to_numeric(numerical_features['run_metrics.val_eval_f1'],errors='coerce')

# Calculate correlations with 'val_eval_f1'
correlations = numerical_features.corr()['run_metrics.val_eval_f1']

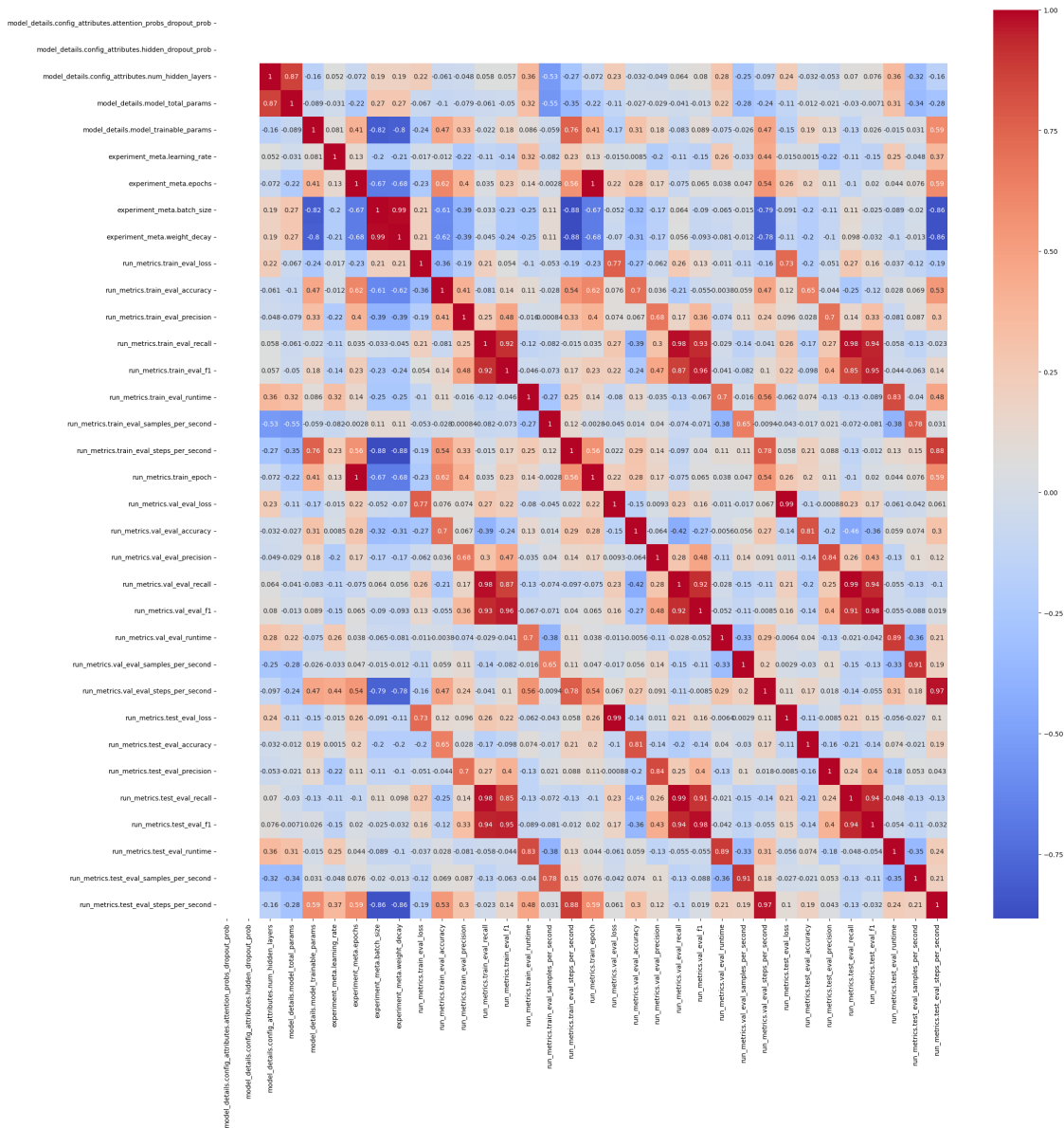
# Print or display the correlations
print(correlations)
plt.figure(figsize=(25, 25))
sns.heatmap(numerical_features.corr(), annot=True, cmap='coolwarm')
plt.show()
```

| | |
|--|-----------|
| model_details.config_attributes.attention_probs_dropout_prob | NaN |
| model_details.config_attributes.hidden_dropout_prob | NaN |
| model_details.config_attributes.num_hidden_layers | 0.079552 |
| model_details.model_total_params | -0.013218 |
| model_details.model_trainable_params | 0.088725 |
| experiment_meta.learning_rate | -0.146438 |
| experiment_meta.epochs | 0.065470 |
| experiment_meta.batch_size | -0.090322 |
| experiment_meta.weight_decay | -0.093220 |
| run_metrics.train_eval_loss | 0.134146 |
| run_metrics.train_eval_accuracy | -0.055486 |
| run_metrics.train_eval_precision | 0.364859 |
| run_metrics.train_eval_recall | 0.928531 |
| run_metrics.train_eval_f1 | 0.962764 |
| run_metrics.train_eval_runtime | -0.066631 |
| run_metrics.train_eval_samples_per_second | -0.071153 |
| run_metrics.train_eval_steps_per_second | 0.039724 |
| run_metrics.train_epoch | 0.065470 |
| run_metrics.val_eval_loss | 0.163439 |
| run_metrics.val_eval_accuracy | -0.272217 |
| run_metrics.val_eval_precision | 0.476833 |
| run_metrics.val_eval_recall | 0.924557 |
| run_metrics.val_eval_f1 | 1.000000 |
| run_metrics.val_eval_runtime | -0.052096 |
| run_metrics.val_eval_samples_per_second | -0.105456 |
| run_metrics.val_eval_steps_per_second | -0.008481 |
| run_metrics.test_eval_loss | 0.155603 |
| run_metrics.test_eval_accuracy | -0.136335 |
| run_metrics.test_eval_precision | 0.396096 |
| run_metrics.test_eval_recall | 0.906551 |

```

run_metrics.test_eval_f1                                0.984679
run_metrics.test_eval_runtime                          -0.055204
run_metrics.test_eval_samples_per_second               -0.087891
run_metrics.test_eval_steps_per_second                 0.019270
Name: run_metrics.val_eval_f1, dtype: float64

```



```

[21]: import pandas as pd

# Select only float columns and 'test_eval_f1'
numerical_features = result_df.select_dtypes(include=['number'])

```

```

# if test_eval_f1 is not numerical for some reason convert to it.
if not pd.api.types.is_numeric_dtype(numerical_features['run_metrics.
↳test_eval_f1']):
    numerical_features['run_metrics.test_eval_f1'] = pd.
↳to_numeric(numerical_features['run_metrics.test_eval_f1'],errors='coerce')

# Calculate correlations with 'test_eval_f1'
correlations = numerical_features.corr()['run_metrics.test_eval_f1']

# Print or display the correlations
print(correlations)

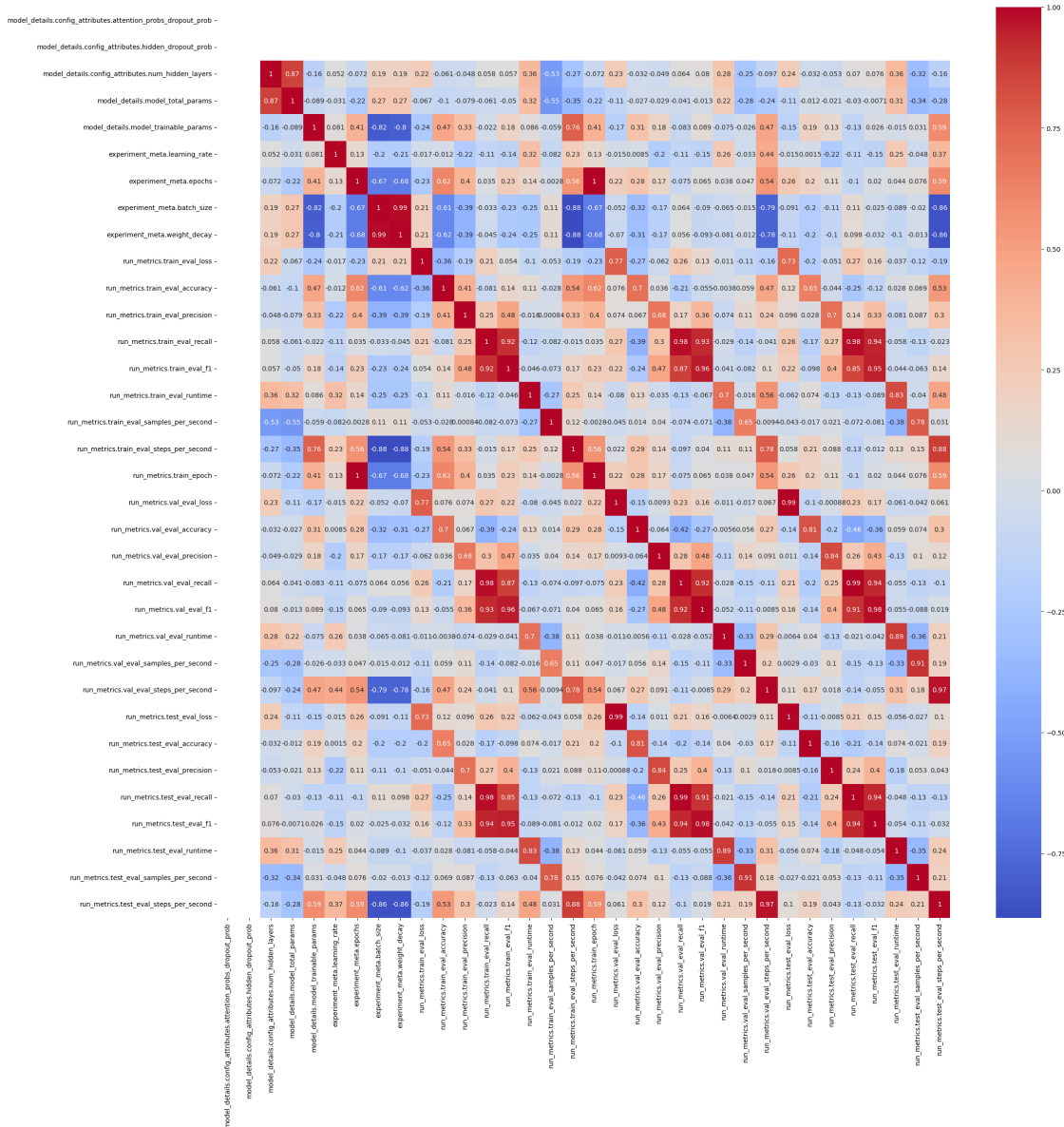
```

```

model_details.config_attributes.attention_probs_dropout_prob      NaN
model_details.config_attributes.hidden_dropout_prob              NaN
model_details.config_attributes.num_hidden_layers                0.076096
model_details.model_total_params                                -0.007088
model_details.model_trainable_params                            0.025648
experiment_meta.learning_rate                                    -0.147384
experiment_meta.epochs                                           0.020381
experiment_meta.batch_size                                       -0.025379
experiment_meta.weight_decay                                     -0.031596
run_metrics.train_eval_loss                                      0.155225
run_metrics.train_eval_accuracy                                 -0.119673
run_metrics.train_eval_precision                                0.328731
run_metrics.train_eval_recall                                    0.941234
run_metrics.train_eval_f1                                        0.949814
run_metrics.train_eval_runtime                                   -0.089306
run_metrics.train_eval_samples_per_second                       -0.081162
run_metrics.train_eval_steps_per_second                         -0.012448
run_metrics.train_epoch                                          0.020381
run_metrics.val_eval_loss                                        0.169496
run_metrics.val_eval_accuracy                                   -0.359808
run_metrics.val_eval_precision                                  0.431582
run_metrics.val_eval_recall                                      0.939597
run_metrics.val_eval_f1                                          0.984679
run_metrics.val_eval_runtime                                    -0.041812
run_metrics.val_eval_samples_per_second                        -0.126446
run_metrics.val_eval_steps_per_second                           -0.054756
run_metrics.test_eval_loss                                       0.154791
run_metrics.test_eval_accuracy                                  -0.144870
run_metrics.test_eval_precision                                  0.403783
run_metrics.test_eval_recall                                      0.936807
run_metrics.test_eval_f1                                         1.000000
run_metrics.test_eval_runtime                                    -0.053893
run_metrics.test_eval_samples_per_second                       -0.111508
run_metrics.test_eval_steps_per_second                          -0.032263
Name: run_metrics.test_eval_f1, dtype: float64

```

```
[22]: import seaborn as sns
import matplotlib.pyplot as plt
plt.figure(figsize=(25, 25))
sns.heatmap(numerical_features.corr(), annot=True, cmap='coolwarm')
plt.show()
```



```
[23]: log_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 248 entries, 0 to 247
Columns: 180 entries, timestamp to
```

```
model_details.config_attributes.task_specific_params.text-generation.max_length
dtypes: bool(17), float64(82), int64(17), object(64)
memory usage: 320.1+ KB
```

[23]:

