COMP0173-CW2-NOTEBOOK-P5

December 11, 2025

1 COMP0173: Coursework 2

The paper HEARTS: A Holistic Framework for Explainable, Sustainable, and Robust Text Stereotype Detection by Theo King, Zekun Wu et al. (2024) presents a comprehensive approach to analysing and detecting stereotypes in text [1]. The authors introduce the HEARTS framework, which integrates model explainability, carbon-efficient training, and accurate evaluation across multiple bias-sensitive datasets. By using transformer-based models such as ALBERT-V2, BERT, and DistilBERT, this research project demonstrates that stereotype detection performance varies significantly across dataset sources, underlining the need for diverse evaluation benchmarks. The paper provides publicly available datasets and code [2], allowing full reproducibility and offering a standardised methodology for future research on bias and stereotype detection in Natural Language Processing (NLP).

While the HEARTS framework evaluates stereotype detection in English, this project adapts the methodology to the Russian context. Russian stereotypes often rely on grammatical gender, morphology, and culture specific tropes. Although Russian is not classified as a low-resource language and many high-performing NLP models are available, there is currently no publicly accessible model specifically designed to detect stereotypes in Russian language. Existing models detecting toxicity or sentiment identify stereotypical and biased sentences only when they include specific patterns, such as insults, slurs, or identity-specific hate speech [8].

To address this gap, I introduce two fine-tuned classifiers, `AI-Forever-RuBert` [10] and `XML-RoBERTa` [11] trained on datasets `RBSA`, and `RBS`, respectively. Understanding these patterns is essential for applications such as content moderation, ensuring the safety of Russian-language LLMs, and monitoring harmful narratives across demographic groups and underrepresented societies. Adapting the HEARTS framework to this new sociolinguistic context illustrates its transferability beyond the English-speaking context and enables a more culturally grounded approach to bias detection, thereby promoting SDG 5: Gender Equality, SDG 10: Reduced Inequalities, and SDG 16: Peace, Justice, and Strong Institutions [5].

2 Instructions

All figures produced during this notebook are stored in the project's `COMP0173_Figures` directory. The corresponding LaTeX-formatted performance comparison tables, jupyter notebooks are stored in `/COMP0173_PDF`. The compiled document are available as `COMP0173-CW2-TABLES.pdf` and `COMP0173_PDF/COMP0173-CW2-NOTEBOOK-XX.pdf`. All prompts used for data augmentation are stored in `COMP0173_Prompts` and the manually collected stereotypes (with English translations) are provided in `COMP0173_Stereotypes`. The datasets used for model training and evaluation are stored in `COMP0173_Data` which contains:

- rubias.tsv — RuBias dataset [6, 7]
- ruster.csv — RuSter dataset (see Part 2 of the notebook for source websites)
- rubist.csv — RBS dataset: RuBias + RuSter augmented with LLM-generated samples (Claude Sonnet), using a zero-shot prompt with examples
- rubist_second.csv — RBSA dataset: RuBias + RuSter augmented with LLM-generated samples using a second prompt version without examples

The notebooks `COMP0173_PDF/COMP0173-CW2-NOTEBOOK-P3.pdf` and `COMP0173_PDF/COMP0173-CW2-NOTEBOOK-P5.pdf` are replications of `COMP0173_PDF/COMP0173-CW2-NOTEBOOK-P2.pdf` and `COMP0173_PDF/COMP0173-CW2-NOTEBOOK-P4.pdf`, where P2 provides the new `RBSA` with second prompt (without examples) and P5 demonstrates the model running ON GPU (the results saved are from GPU fine-tuning).

3 Technical Implementation (70%)

[1]:
```
# %%capture
# pip install -r requirements.txt
# pip install transformers
# pip install --upgrade transformers
# pip install --upgrade tokenizers
# pip install -U sentence-transformers
# pip install natasha
# pip install datasets
# pip install --user -U nltk
# conda install -c anaconda nltk
# pip install --upgrade openai pandas tqdm
# pip install dotenv
```

[2]:
```
# pip install -U pip setuptools wheel
# pip install -U spacy
# python -m spacy download en_core_web_trf
# python -m spacy download en_core_web_sm
# python -m spacy download ru_core_news_lg

# # GPU
# pip install -U 'spacy[cuda12x]'
# # GPU - Train Models
# pip install -U 'spacy[cuda12x,transformers,lookups]'
```

[3]:
```
# Import the libraries
import random, numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set(color_codes=True)
plt.style.use('seaborn-v0_8')
```

```
# To ignore warnings
import warnings
warnings.filterwarnings('ignore')
np.random.seed(23)

warnings.filterwarnings(
    "ignore",
    message="pkg_resources is deprecated as an API"
)
```

[4]:
```
# Import libraries
import pandas as pd
import os
import sys
import importlib.util, pathlib
from pathlib import Path
import warnings
from importlib import reload
from importlib.machinery import SourceFileLoader
from IPython.display import display
import pandas as pd
from pathlib import Path
import re
import difflib
import string
from collections import defaultdict
import json
```

[5]:
```
import torch
import transformers
from transformers import AutoModelForMaskedLM, XLMWithLMHeadModel
from transformers import AutoTokenizer, AutoConfig
from transformers import TrainingArguments, Trainer
from sentence_transformers import SentenceTransformer, util
import platform
from datasets import Dataset
# import spacy
import requests
from tqdm import tqdm
import yaml
```

[6]:
```
sys.path.append("Exploratory Data Analysis")
sys.path.append("Model Training and Evaluation")
```

[7]:
```
# Check the GPU host (UCL access)
print("CUDA available:", torch.cuda.is_available())
print("Device:", torch.cuda.get_device_name(0))
```

```
# # Path
# import os
# os.chdir("/tmp/HEARTS-Text-Stereotype-Detection")
# os.getcwd()
```

```
CUDA available: True
Device: Tesla T4
```

3.1 Part 4: Adapt the model architecture and training pipeline to your local context

3.1.1 *Question 1 :* Justify architectural modifications for new context

To adapt the HEARTS framework to the Russian context, I kept the original fine-tuning pipeline while substituting ALBERT-V2, BERT, and DistilBERT with encoder models optimised for Russian text. The specific models I fine-tuned include:

- DeepPavlov/RuBERT [9]
- AI-Forever/RuBERT [10]
- XLM-RoBERTa (multilingual) [11]
- Logistic Regression baselines using TF-IDF and SpaCy embeddings.

Each model was configured as a binary stereotype classifier and trained separately on the RBS and RBSA datasets using the Hugging Face AutoModelForSequenceClassification architecture, with an 80/20 train-test split. To support sustainability goals, CodeCarbon was integrated into the pipeline to monitor emissions during fine-tuning. All models finished training in under 10 minutes per dataset, with total estimated emissions of less than 2 grams of CO for each run.

Table 5: Fine-tuned RuBert Model: Hyperparameter Choices and Training Setup

Parameter	**Value**
Batch Size	64
Learning Rate	2×10^{-5}
Epochs	6
Training Device	GPU (NVIDIA Tesla T4)
Approximate Runtime	< 10 Minutes

Table 6: Fine-tuned AI-Forever/RuBert Model: Configuration

Category	Details
Key Information	
Model Name	stereotype_bias_classifier_rubert
Base Architecture	BertForSequenceClassification
Number of Parameters	178,308,866
Vocabulary Size	120,138
Labels	{0, 1}
Model Configuration and Capacity	
Embedding Dimensionality	768
Intermediate Layer Size	3072
Hidden Layer Size	768
Number of Hidden Layers	12
Number of Attention Heads	12
Regularisation Hyperparameters	
Hidden Layer Activation	gelu
Hidden Layer Dropout Probability	0.1
Attention Head Dropout Probability	0.1
Classification Layer Dropout Probability	None
Layer Normalisation Epsilon	1.0×10^{-12}

[8]:
```
# Load final version
rubist = pd.read_csv("COMP0173_Data/rubist.csv", encoding="utf-8")
rubist_second = pd.read_csv("COMP0173_Data/rubist_second.csv", encoding="utf-8")
```

Train models

[9]:
```
import os
os.environ["HF_HOME"] = "/tmp/hf"
os.environ["TRANSFORMERS_CACHE"] = "/tmp/hf"
os.makedirs("/tmp/hf", exist_ok=True)
import gc
```

[10]:
```
from Logistic_Regression_Russian import (data_loader, train_model,␣
 ↪evaluate_model)

gc.collect()
torch.cuda.empty_cache()

# Load and combine relevant datasets
train_data_rubist, test_data_rubist = data_loader(csv_file_path='COMP0173_Data/
 ↪rubist.csv', labelling_criteria='stereotype', dataset_name='rubist',␣
 ↪sample_size=1000000, num_examples=5)
train_data_rubist_second, test_data_rubist_second =␣
 ↪data_loader(csv_file_path='COMP0173_Data/rubist_second.csv',␣
 ↪labelling_criteria='stereotype', dataset_name='rubist_second',␣
 ↪sample_size=1000000, num_examples=5)
```

```
# Execute full pipeline for logistic regression tfidf model
train_model(train_data_rubist, model_output_base_dir='model_output_LR_tfidf',␣
 ↪dataset_name='rubist_trained', feature_type='tfidf', seed=42)
evaluate_model(test_data_rubist, model_output_dir='model_output_LR_tfidf/
 ↪rubist_trained', result_output_base_dir='result_output_LR_tfidf',␣
 ↪dataset_name='rubist', feature_type='tfidf', seed=42)

gc.collect()
torch.cuda.empty_cache()

train_model(train_data_rubist_second,␣
 ↪model_output_base_dir='model_output_LR_tfidf',␣
 ↪dataset_name='rubist_second_trained', feature_type='tfidf', seed=42)
evaluate_model(test_data_rubist_second, model_output_dir='model_output_LR_tfidf/
 ↪rubist_second_trained', result_output_base_dir='result_output_LR_tfidf',␣
 ↪dataset_name='rubist_second', feature_type='tfidf', seed=42)

gc.collect()
torch.cuda.empty_cache()

# Execute full pipeline for logistic regression embedding model
train_model(train_data_rubist,␣
 ↪model_output_base_dir='model_output_LR_embedding',␣
 ↪dataset_name='rubist_trained', feature_type='embedding', seed=42)
evaluate_model(test_data_rubist, model_output_dir='model_output_LR_embedding/
 ↪rubist_trained', result_output_base_dir='result_output_LR_embedding',␣
 ↪dataset_name='rubist', feature_type='embedding', seed=42)

gc.collect()
torch.cuda.empty_cache()

train_model(train_data_rubist_second,␣
 ↪model_output_base_dir='model_output_LR_embedding',␣
 ↪dataset_name='rubist_second_trained', feature_type='embedding', seed=42)
evaluate_model(test_data_rubist_second,␣
 ↪model_output_dir='model_output_LR_embedding/rubist_second_trained',␣
 ↪result_output_base_dir='result_output_LR_embedding',␣
 ↪dataset_name='rubist_second', feature_type='embedding', seed=42)
```

```
First few examples from the training data:
     stereotype_type                                              text  \
168       profession
2883     nationality                                …
2919           lgbtq                                …
3412      profession                                …
```

```
732           lgbtq

     category data_name
168          1    rubist
2883         0    rubist
2919         1    rubist
3412         0    rubist
732          1    rubist
First few examples from the testing data:
     stereotype_type                                                text  \
2726      profession                                …
1066          gender                               …
2547     nationality                               …
1834      profession                               …
1825     nationality                               …

     category data_name
2726         0    rubist
1066         0    rubist
2547         0    rubist
1834         0    rubist
1825         0    rubist
Train data size: 3372
Test data size: 844
First few examples from the training data:
     stereotype_type                                          text  category  \
1005          gender                                            1
1001          gender                                          0
2619     nationality                                           0
2213          gender                                     0
934           gender                                      0

         data_name
1005  rubist_second
1001  rubist_second
2619  rubist_second
2213  rubist_second
934   rubist_second
First few examples from the testing data:
     stereotype_type                                                text  \
80        profession
1814          gender
2277      profession                                …
61        profession                                …
755           gender                               …

     category      data_name
80          0  rubist_second
```

```
1814         0  rubist_second
2277         0  rubist_second
61           0  rubist_second
755          0  rubist_second
Train data size: 2336
Test data size: 584
Number of unique labels: 2
Testing C=0.01, penalty=l1 => F1 Score: 0.39785905441570024
Testing C=0.01, penalty=l2 => F1 Score: 0.39785905441570024
Testing C=0.01, penalty=None => F1 Score: 0.9466666666666667
Testing C=0.1, penalty=l1 => F1 Score: 0.6326957285464712
Testing C=0.1, penalty=l2 => F1 Score: 0.6120484633882092
Testing C=0.1, penalty=None => F1 Score: 0.9466666666666667
Testing C=1, penalty=l1 => F1 Score: 0.8929148318366249
Testing C=1, penalty=l2 => F1 Score: 0.8824924844557498
Testing C=1, penalty=None => F1 Score: 0.9466666666666667
Best model parameters: {'C': 0.01, 'penalty': None}
Model and vectorizer saved to model_output_LR_tfidf/rubist_trained
Estimated total emissions: 1.3855943050266042e-05 kg CO2
Number of unique labels: 2
Number of unique labels: 2
Testing C=0.01, penalty=l1 => F1 Score: 0.4
Testing C=0.01, penalty=l2 => F1 Score: 0.4
Testing C=0.01, penalty=None => F1 Score: 0.5407413941385317
Testing C=0.1, penalty=l1 => F1 Score: 0.4
Testing C=0.1, penalty=l2 => F1 Score: 0.4
Testing C=0.1, penalty=None => F1 Score: 0.5407413941385317
Testing C=1, penalty=l1 => F1 Score: 0.46003885576472164
Testing C=1, penalty=l2 => F1 Score: 0.42689732142857145
Testing C=1, penalty=None => F1 Score: 0.5407413941385317
Best model parameters: {'C': 0.01, 'penalty': None}
Model and vectorizer saved to model_output_LR_tfidf/rubist_second_trained
Estimated total emissions: 1.2133621719059336e-05 kg CO2
Number of unique labels: 2
Number of unique labels: 2

Computing embeddings: 100%|      | 3372/3372 [00:33<00:00, 100.47it/s]

Testing C=0.01, penalty=l1 => F1 Score: 0.39785905441570024
Testing C=0.01, penalty=l2 => F1 Score: 0.8377403846153846
Testing C=0.01, penalty=None => F1 Score: 0.9196396682408032
Testing C=0.1, penalty=l1 => F1 Score: 0.8833912566306933
Testing C=0.1, penalty=l2 => F1 Score: 0.894211324570273
Testing C=0.1, penalty=None => F1 Score: 0.9196396682408032
Testing C=1, penalty=l1 => F1 Score: 0.9212241604072258
Testing C=1, penalty=l2 => F1 Score: 0.9172978203631145
Testing C=1, penalty=None => F1 Score: 0.9196396682408032
Best model parameters: {'C': 1, 'penalty': 'l1'}
Model and vectorizer saved to model_output_LR_embedding/rubist_trained
```

```
Estimated total emissions: 0.0001950465552109412 kg CO2
Number of unique labels: 2

Computing embeddings: 100%|      | 844/844 [00:08<00:00, 100.36it/s]

Number of unique labels: 2

Computing embeddings: 100%|      | 2336/2336 [00:20<00:00, 112.08it/s]

Testing C=0.01, penalty=l1 => F1 Score: 0.4
Testing C=0.01, penalty=l2 => F1 Score: 0.4068829055705911
Testing C=0.01, penalty=None => F1 Score: 0.627725258253562
Testing C=0.1, penalty=l1 => F1 Score: 0.4052597071464996
Testing C=0.1, penalty=l2 => F1 Score: 0.5424300867888139
Testing C=0.1, penalty=None => F1 Score: 0.627725258253562
Testing C=1, penalty=l1 => F1 Score: 0.6112426035502958
Testing C=1, penalty=l2 => F1 Score: 0.6005917159763313
Testing C=1, penalty=None => F1 Score: 0.627725258253562
Best model parameters: {'C': 0.01, 'penalty': None}
Model and vectorizer saved to model_output_LR_embedding/rubist_second_trained
Estimated total emissions: 0.00013900726188074115 kg CO2
Number of unique labels: 2

Computing embeddings: 100%|      | 584/584 [00:05<00:00, 112.28it/s]
```

[10]:

	precision	recall	f1-score	support
0	0.763547	0.796915	0.779874	389.000000
1	0.556180	0.507692	0.530831	195.000000
accuracy	0.700342	0.700342	0.700342	0.700342
macro avg	0.659863	0.652304	0.655353	584.000000
weighted avg	0.694306	0.700342	0.696718	584.000000

3.1.2 *Question 2 :* Document hyperparameter tuning process - GPU

Hyperparameter tuning followed the structure of the original HEARTS pipeline but was adapted to Russian-language models and the two augmented datasets (RBS and RBSA). All experiments were run on a GPU-enabled environment to support efficient fine-tuning of transformer models. Before each run, GPU memory was cleared using:

`gc.collect() torch.cuda.empty_cache()`

The tuning process began by loading the two datasets (rubist.csv and rubist_second.csv) using the customised data_loader() function. For each model, a consistent training configuration was used to enable fair comparison. Unfortunately due to disk quota, it was not possible to run the `XLM-Roberta` model on GPU machine.

[12]:
```
from BERT_Models_Fine_Tuning_Russian import (data_loader, train_model,␣
 ↪evaluate_model)

gc.collect()
torch.cuda.empty_cache()
```

```
# Load and combine relevant datasets
train_data_rubist, test_data_rubist = data_loader(csv_file_path='COMP0173_Data/
 ↪rubist.csv', labelling_criteria='stereotype', dataset_name='rubist',␣
 ↪sample_size=1000000, num_examples=5)
train_data_rubist_second, test_data_rubist_second =␣
 ↪data_loader(csv_file_path='COMP0173_Data/rubist_second.csv',␣
 ↪labelling_criteria='stereotype', dataset_name='rubist_second',␣
 ↪sample_size=1000000, num_examples=5)

# Execute full pipeline for Deepavlov model
train_model(train_data_rubist, model_path='DeepPavlov/rubert-base-cased',␣
 ↪batch_size=64, epoch=6, learning_rate=2e-5,␣
 ↪model_output_base_dir='model_output_deeppavlov_rubert',␣
 ↪dataset_name='rubist_trained', seed=42)
evaluate_model(test_data_rubist,␣
 ↪model_output_dir='model_output_deeppavlov_rubert/rubist_trained',␣
 ↪result_output_base_dir='result_output_deeppavlov_rubert',␣
 ↪dataset_name='rubist_trained', seed=42)

gc.collect()
torch.cuda.empty_cache()

train_model(train_data_rubist_second, model_path='DeepPavlov/
 ↪rubert-base-cased', batch_size=64, epoch=6, learning_rate=2e-5,␣
 ↪model_output_base_dir='model_output_deeppavlov_rubert',␣
 ↪dataset_name='rubist_second_trained', seed=42)
evaluate_model(test_data_rubist_second,␣
 ↪model_output_dir='model_output_deeppavlov_rubert/rubist_second_trained',␣
 ↪result_output_base_dir='result_output_deeppavlov_rubert',␣
 ↪dataset_name='rubist_second_trained', seed=42)
```

```
First few examples from the training data:
     stereotype_type                                                text  \
168       profession
2883     nationality                                 …
2919           lgbtq                                 …
3412      profession                                 …
732            lgbtq

      category data_name
168          1    rubist
2883         0    rubist
2919         1    rubist
3412         0    rubist
732          1    rubist
First few examples from the testing data:
```

```
     stereotype_type                                                  text  \
2726      profession                                …
1066          gender                                …
2547     nationality                                …
1834      profession                                …
1825     nationality                                …

      category data_name
2726         0    rubist
1066         0    rubist
2547         0    rubist
1834         0    rubist
1825         0    rubist
Train data size: 3372
Test data size: 844
First few examples from the training data:
     stereotype_type                                          text  category  \
1005          gender                                            1
1001          gender                                          0
2619     nationality                                           0
2213          gender                                     0
934           gender                                      0

          data_name
1005  rubist_second
1001  rubist_second
2619  rubist_second
2213  rubist_second
934   rubist_second
First few examples from the testing data:
     stereotype_type                                                  text  \
80        profession
1814          gender
2277      profession                                …
61        profession                                …
755           gender                                …

      category      data_name
80           0  rubist_second
1814         0  rubist_second
2277         0  rubist_second
61           0  rubist_second
755          0  rubist_second
Train data size: 2336
Test data size: 584
Number of unique labels: 2

loading configuration file config.json from cache at
```

```
/home/ec2-user/.cache/huggingface/hub/models--DeepPavlov--rubert-base-
cased/snapshots/4036cab694767a299f2b9e6492909664d9414229/config.json
Model config BertConfig {
  "_name_or_path": "DeepPavlov/rubert-base-cased",
  "architectures": [
    "BertModel"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_past": true,
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 119547
}

loading weights file pytorch_model.bin from cache at
/home/ec2-user/.cache/huggingface/hub/models--DeepPavlov--rubert-base-
cased/snapshots/4036cab694767a299f2b9e6492909664d9414229/pytorch_model.bin
Attempting to create safetensors variant
Attempting to convert .bin model on the fly to safetensors.
Some weights of the model checkpoint at DeepPavlov/rubert-base-cased were not
used when initializing BertForSequenceClassification: ['cls.predictions.bias',
'cls.predictions.decoder.bias', 'cls.predictions.decoder.weight',
'cls.predictions.transform.LayerNorm.bias',
'cls.predictions.transform.LayerNorm.weight',
'cls.predictions.transform.dense.bias',
'cls.predictions.transform.dense.weight', 'cls.seq_relationship.bias',
'cls.seq_relationship.weight']
- This IS expected if you are initializing BertForSequenceClassification from
```

```
the checkpoint of a model trained on another task or with another architecture
(e.g. initializing a BertForSequenceClassification model from a
BertForPreTraining model).
- This IS NOT expected if you are initializing BertForSequenceClassification
from the checkpoint of a model that you expect to be exactly identical
(initializing a BertForSequenceClassification model from a
BertForSequenceClassification model).
Some weights of BertForSequenceClassification were not initialized from the
model checkpoint at DeepPavlov/rubert-base-cased and are newly initialized:
['classifier.bias', 'classifier.weight']
You should probably TRAIN this model on a down-stream task to be able to use it
for predictions and inference.
loading configuration file config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--DeepPavlov--rubert-base-
cased/snapshots/4036cab694767a299f2b9e6492909664d9414229/config.json
Model config BertConfig {
  "_name_or_path": "DeepPavlov/rubert-base-cased",
  "architectures": [
    "BertModel"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_past": true,
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 119547
}

loading file vocab.txt from cache at
```

```
/home/ec2-user/.cache/huggingface/hub/models--DeepPavlov--rubert-base-
cased/snapshots/4036cab694767a299f2b9e6492909664d9414229/vocab.txt
loading file tokenizer.json from cache at None
loading file added_tokens.json from cache at None
loading file special_tokens_map.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--DeepPavlov--rubert-base-
cased/snapshots/4036cab694767a299f2b9e6492909664d9414229/special_tokens_map.json
loading file tokenizer_config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--DeepPavlov--rubert-base-
cased/snapshots/4036cab694767a299f2b9e6492909664d9414229/tokenizer_config.json
loading configuration file config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--DeepPavlov--rubert-base-
cased/snapshots/4036cab694767a299f2b9e6492909664d9414229/config.json
Model config BertConfig {
  "_name_or_path": "DeepPavlov/rubert-base-cased",
  "architectures": [
    "BertModel"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_past": true,
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 119547
}

loading configuration file config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--DeepPavlov--rubert-base-
cased/snapshots/4036cab694767a299f2b9e6492909664d9414229/config.json
```

```
Model config BertConfig {
  "_name_or_path": "DeepPavlov/rubert-base-cased",
  "architectures": [
    "BertModel"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_past": true,
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 119547
}


Map:   0%|          | 0/2697 [00:00<?, ? examples/s]

Map:   0%|          | 0/2697 [00:00<?, ? examples/s]

Sample tokenized input from train: {'stereotype_type': 'gender', 'text':
'                                      ',
'category': 0, 'data_name': 'rubist', '__index_level_0__': 1317, 'input_ids':
[101, 12528, 23558, 29466, 12938, 21264, 57041, 869, 1516, 33165, 27585, 9210,
102, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], 'token_type_ids': [0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
'attention_mask': [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0], 'labels': 0}

Map:   0%|          | 0/675 [00:00<?, ? examples/s]

Map:   0%|          | 0/675 [00:00<?, ? examples/s]

PyTorch: setting up devices
```

```
Sample tokenized input from validation: {'stereotype_type': 'gender', 'text':
'                                       ',
'category': 0, 'data_name': 'rubist', '__index_level_0__': 1317, 'input_ids':
[101, 12528, 23558, 29466, 12938, 21264, 57041, 869, 1516, 33165, 27585, 9210,
102, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], 'token_type_ids': [0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
'attention_mask': [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0], 'labels': 0}

The default value for the training argument `--report_to` will change in v5
(from all installed integrations to none). In v5, you will need to use
`--report_to all` to get the same behavior as now. You should start updating
your code and make this info disappear :-).
The following columns in the training set don't have a corresponding argument in
`BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.
***** Running training *****
  Num examples = 2,697
  Num Epochs = 6
  Instantaneous batch size per device = 64
  Total train batch size (w. parallel, distributed & accumulation) = 64
  Gradient Accumulation steps = 1
  Total optimization steps = 258
  Number of trainable parameters = 177,854,978

<IPython.core.display.HTML object>

The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 675
  Batch size = 64
Saving model checkpoint to
model_output_deeppavlov_rubert/rubist_trained/checkpoint-43
Configuration saved in
model_output_deeppavlov_rubert/rubist_trained/checkpoint-43/config.json
Model weights saved in
model_output_deeppavlov_rubert/rubist_trained/checkpoint-43/model.safetensors
tokenizer config file saved in model_output_deeppavlov_rubert/rubist_trained/che
ckpoint-43/tokenizer_config.json
Special tokens file saved in model_output_deeppavlov_rubert/rubist_trained/check
point-43/special_tokens_map.json
The following columns in the evaluation set don't have a corresponding argument
```

```
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 675
  Batch size = 64
Saving model checkpoint to
model_output_deeppavlov_rubert/rubist_trained/checkpoint-86
Configuration saved in
model_output_deeppavlov_rubert/rubist_trained/checkpoint-86/config.json
Model weights saved in
model_output_deeppavlov_rubert/rubist_trained/checkpoint-86/model.safetensors
tokenizer config file saved in model_output_deeppavlov_rubert/rubist_trained/che
ckpoint-86/tokenizer_config.json
Special tokens file saved in model_output_deeppavlov_rubert/rubist_trained/check
point-86/special_tokens_map.json
Deleting older checkpoint
[model_output_deeppavlov_rubert/rubist_trained/checkpoint-43] due to
args.save_total_limit
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 675
  Batch size = 64
Saving model checkpoint to
model_output_deeppavlov_rubert/rubist_trained/checkpoint-129
Configuration saved in
model_output_deeppavlov_rubert/rubist_trained/checkpoint-129/config.json
Model weights saved in
model_output_deeppavlov_rubert/rubist_trained/checkpoint-129/model.safetensors
tokenizer config file saved in model_output_deeppavlov_rubert/rubist_trained/che
ckpoint-129/tokenizer_config.json
Special tokens file saved in model_output_deeppavlov_rubert/rubist_trained/check
point-129/special_tokens_map.json
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 675
```

```
  Batch size = 64
Saving model checkpoint to
model_output_deeppavlov_rubert/rubist_trained/checkpoint-172
Configuration saved in
model_output_deeppavlov_rubert/rubist_trained/checkpoint-172/config.json
Model weights saved in
model_output_deeppavlov_rubert/rubist_trained/checkpoint-172/model.safetensors
tokenizer config file saved in model_output_deeppavlov_rubert/rubist_trained/che
ckpoint-172/tokenizer_config.json
Special tokens file saved in model_output_deeppavlov_rubert/rubist_trained/check
point-172/special_tokens_map.json
Deleting older checkpoint
[model_output_deeppavlov_rubert/rubist_trained/checkpoint-86] due to
args.save_total_limit
Deleting older checkpoint
[model_output_deeppavlov_rubert/rubist_trained/checkpoint-129] due to
args.save_total_limit
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 675
  Batch size = 64
Saving model checkpoint to
model_output_deeppavlov_rubert/rubist_trained/checkpoint-215
Configuration saved in
model_output_deeppavlov_rubert/rubist_trained/checkpoint-215/config.json
Model weights saved in
model_output_deeppavlov_rubert/rubist_trained/checkpoint-215/model.safetensors
tokenizer config file saved in model_output_deeppavlov_rubert/rubist_trained/che
ckpoint-215/tokenizer_config.json
Special tokens file saved in model_output_deeppavlov_rubert/rubist_trained/check
point-215/special_tokens_map.json
Saving model checkpoint to
model_output_deeppavlov_rubert/rubist_trained/checkpoint-258
Configuration saved in
model_output_deeppavlov_rubert/rubist_trained/checkpoint-258/config.json
Model weights saved in
model_output_deeppavlov_rubert/rubist_trained/checkpoint-258/model.safetensors
tokenizer config file saved in model_output_deeppavlov_rubert/rubist_trained/che
ckpoint-258/tokenizer_config.json
Special tokens file saved in model_output_deeppavlov_rubert/rubist_trained/check
point-258/special_tokens_map.json
Deleting older checkpoint
[model_output_deeppavlov_rubert/rubist_trained/checkpoint-215] due to
```

```
args.save_total_limit
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 675
  Batch size = 64
Saving model checkpoint to
model_output_deeppavlov_rubert/rubist_trained/checkpoint-258
Configuration saved in
model_output_deeppavlov_rubert/rubist_trained/checkpoint-258/config.json
Model weights saved in
model_output_deeppavlov_rubert/rubist_trained/checkpoint-258/model.safetensors
tokenizer config file saved in model_output_deeppavlov_rubert/rubist_trained/che
ckpoint-258/tokenizer_config.json
Special tokens file saved in model_output_deeppavlov_rubert/rubist_trained/check
point-258/special_tokens_map.json


Training completed. Do not forget to share your model on huggingface.co/models
=)


Loading best model from
model_output_deeppavlov_rubert/rubist_trained/checkpoint-172 (score:
0.0698399767279625).
Deleting older checkpoint
[model_output_deeppavlov_rubert/rubist_trained/checkpoint-258] due to
args.save_total_limit
Saving model checkpoint to model_output_deeppavlov_rubert/rubist_trained
Configuration saved in model_output_deeppavlov_rubert/rubist_trained/config.json
Model weights saved in
model_output_deeppavlov_rubert/rubist_trained/model.safetensors
tokenizer config file saved in
model_output_deeppavlov_rubert/rubist_trained/tokenizer_config.json
Special tokens file saved in
model_output_deeppavlov_rubert/rubist_trained/special_tokens_map.json
loading configuration file
model_output_deeppavlov_rubert/rubist_trained/config.json
Model config BertConfig {
  "_name_or_path": "model_output_deeppavlov_rubert/rubist_trained",
  "architectures": [
    "BertForSequenceClassification"
  ],
  "attention_probs_dropout_prob": 0.1,
```

```
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_past": true,
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "problem_type": "single_label_classification",
  "torch_dtype": "float32",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 119547
}
```


```
loading weights file
model_output_deeppavlov_rubert/rubist_trained/model.safetensors
```

```
Estimated total emissions: 0.0012824234634039179 kg CO2
Number of unique labels: 2
```

```
All model checkpoint weights were used when initializing
BertForSequenceClassification.
```

```
All the weights of BertForSequenceClassification were initialized from the model
checkpoint at model_output_deeppavlov_rubert/rubist_trained.
If your task is similar to the task the model of the checkpoint was trained on,
you can already use BertForSequenceClassification for predictions without
further training.
loading file vocab.txt
loading file tokenizer.json
loading file added_tokens.json
loading file special_tokens_map.json
loading file tokenizer_config.json
```

```
Map:   0%|          | 0/844 [00:00<?, ? examples/s]
```

```
Map:   0%|          | 0/844 [00:00<?, ? examples/s]

Disabling tokenizer parallelism, we're using DataLoader multithreading already

Sample tokenized input from test: {'stereotype_type': 'profession', 'text':
'                                    ', 'category': 0,
'data_name': 'rubist', '__index_level_0__': 2726, 'input_ids': [101, 94818,
11894, 38663, 13904, 2010, 54564, 851, 84652, 868, 10508, 1703, 13158, 23726,
102, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], 'token_type_ids': [0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
'attention_mask': [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0], 'labels': 0}
Number of unique labels: 2

loading configuration file config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--DeepPavlov--rubert-base-
cased/snapshots/4036cab694767a299f2b9e6492909664d9414229/config.json
Model config BertConfig {
  "_name_or_path": "DeepPavlov/rubert-base-cased",
  "architectures": [
    "BertModel"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_past": true,
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 119547
}
```

```
loading weights file pytorch_model.bin from cache at
/home/ec2-user/.cache/huggingface/hub/models--DeepPavlov--rubert-base-
cased/snapshots/4036cab694767a299f2b9e6492909664d9414229/pytorch_model.bin
Attempting to create safetensors variant
Attempting to convert .bin model on the fly to safetensors.
Some weights of the model checkpoint at DeepPavlov/rubert-base-cased were not
used when initializing BertForSequenceClassification: ['cls.predictions.bias',
'cls.predictions.decoder.bias', 'cls.predictions.decoder.weight',
'cls.predictions.transform.LayerNorm.bias',
'cls.predictions.transform.LayerNorm.weight',
'cls.predictions.transform.dense.bias',
'cls.predictions.transform.dense.weight', 'cls.seq_relationship.bias',
'cls.seq_relationship.weight']
- This IS expected if you are initializing BertForSequenceClassification from
the checkpoint of a model trained on another task or with another architecture
(e.g. initializing a BertForSequenceClassification model from a
BertForPreTraining model).
- This IS NOT expected if you are initializing BertForSequenceClassification
from the checkpoint of a model that you expect to be exactly identical
(initializing a BertForSequenceClassification model from a
BertForSequenceClassification model).
Some weights of BertForSequenceClassification were not initialized from the
model checkpoint at DeepPavlov/rubert-base-cased and are newly initialized:
['classifier.bias', 'classifier.weight']
You should probably TRAIN this model on a down-stream task to be able to use it
for predictions and inference.
loading configuration file config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--DeepPavlov--rubert-base-
cased/snapshots/4036cab694767a299f2b9e6492909664d9414229/config.json
Model config BertConfig {
  "_name_or_path": "DeepPavlov/rubert-base-cased",
  "architectures": [
    "BertModel"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_past": true,
```

```
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 119547
}

loading file vocab.txt from cache at
/home/ec2-user/.cache/huggingface/hub/models--DeepPavlov--rubert-base-
cased/snapshots/4036cab694767a299f2b9e6492909664d9414229/vocab.txt
loading file tokenizer.json from cache at None
loading file added_tokens.json from cache at None
loading file special_tokens_map.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--DeepPavlov--rubert-base-
cased/snapshots/4036cab694767a299f2b9e6492909664d9414229/special_tokens_map.json
loading file tokenizer_config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--DeepPavlov--rubert-base-
cased/snapshots/4036cab694767a299f2b9e6492909664d9414229/tokenizer_config.json
loading configuration file config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--DeepPavlov--rubert-base-
cased/snapshots/4036cab694767a299f2b9e6492909664d9414229/config.json
Model config BertConfig {
  "_name_or_path": "DeepPavlov/rubert-base-cased",
  "architectures": [
    "BertModel"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_past": true,
  "pad_token_id": 0,
  "pooler_fc_size": 768,
```

```
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 119547
}

loading configuration file config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--DeepPavlov--rubert-base-
cased/snapshots/4036cab694767a299f2b9e6492909664d9414229/config.json
Model config BertConfig {
  "_name_or_path": "DeepPavlov/rubert-base-cased",
  "architectures": [
    "BertModel"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_past": true,
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 119547
}

Map:   0%|          | 0/1868 [00:00<?, ? examples/s]
```

```
Map:   0%|          | 0/1868 [00:00<?, ? examples/s]

Sample tokenized input from train: {'stereotype_type': 'nationality', 'text':
'                ', 'category': 0, 'data_name': 'rubist_second',
'__index_level_0__': 1567, 'input_ids': [101, 4752, 92660, 4402, 7805, 14741,
23939, 102, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
'token_type_ids': [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0], 'attention_mask': [1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], 'labels': 0}

Map:   0%|          | 0/468 [00:00<?, ? examples/s]

Map:   0%|          | 0/468 [00:00<?, ? examples/s]

PyTorch: setting up devices
The default value for the training argument `--report_to` will change in v5
(from all installed integrations to none). In v5, you will need to use
`--report_to all` to get the same behavior as now. You should start updating
your code and make this info disappear :-).

Sample tokenized input from validation: {'stereotype_type': 'nationality',
'text': '                ', 'category': 0, 'data_name':
'rubist_second', '__index_level_0__': 1567, 'input_ids': [101, 4752, 92660,
4402, 7805, 14741, 23939, 102, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0], 'token_type_ids': [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], 'attention_mask': [1, 1, 1, 1, 1, 1,
1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
'labels': 0}

The following columns in the training set don't have a corresponding argument in
`BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.
***** Running training *****
  Num examples = 1,868
  Num Epochs = 6
  Instantaneous batch size per device = 64
  Total train batch size (w. parallel, distributed & accumulation) = 64
  Gradient Accumulation steps = 1
  Total optimization steps = 180
  Number of trainable parameters = 177,854,978

<IPython.core.display.HTML object>

The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
```

```
  Num examples = 468
  Batch size = 64
Saving model checkpoint to
model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-30
Configuration saved in
model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-30/config.json
Model weights saved in model_output_deeppavlov_rubert/rubist_second_trained/chec
kpoint-30/model.safetensors
tokenizer config file saved in model_output_deeppavlov_rubert/rubist_second_trai
ned/checkpoint-30/tokenizer_config.json
Special tokens file saved in model_output_deeppavlov_rubert/rubist_second_traine
d/checkpoint-30/special_tokens_map.json
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 468
  Batch size = 64
Saving model checkpoint to
model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-60
Configuration saved in
model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-60/config.json
Model weights saved in model_output_deeppavlov_rubert/rubist_second_trained/chec
kpoint-60/model.safetensors
tokenizer config file saved in model_output_deeppavlov_rubert/rubist_second_trai
ned/checkpoint-60/tokenizer_config.json
Special tokens file saved in model_output_deeppavlov_rubert/rubist_second_traine
d/checkpoint-60/special_tokens_map.json
Deleting older checkpoint
[model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-30] due to
args.save_total_limit
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 468
  Batch size = 64
Saving model checkpoint to
model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-90
Configuration saved in
model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-90/config.json
Model weights saved in model_output_deeppavlov_rubert/rubist_second_trained/chec
```

```
kpoint-90/model.safetensors
tokenizer config file saved in model_output_deeppavlov_rubert/rubist_second_trai
ned/checkpoint-90/tokenizer_config.json
Special tokens file saved in model_output_deeppavlov_rubert/rubist_second_traine
d/checkpoint-90/special_tokens_map.json
Deleting older checkpoint
[model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-60] due to
args.save_total_limit
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 468
  Batch size = 64
Saving model checkpoint to
model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-120
Configuration saved in
model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-120/config.json
Model weights saved in model_output_deeppavlov_rubert/rubist_second_trained/chec
kpoint-120/model.safetensors
tokenizer config file saved in model_output_deeppavlov_rubert/rubist_second_trai
ned/checkpoint-120/tokenizer_config.json
Special tokens file saved in model_output_deeppavlov_rubert/rubist_second_traine
d/checkpoint-120/special_tokens_map.json
Deleting older checkpoint
[model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-90] due to
args.save_total_limit
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 468
  Batch size = 64
Saving model checkpoint to
model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-150
Configuration saved in
model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-150/config.json
Model weights saved in model_output_deeppavlov_rubert/rubist_second_trained/chec
kpoint-150/model.safetensors
tokenizer config file saved in model_output_deeppavlov_rubert/rubist_second_trai
ned/checkpoint-150/tokenizer_config.json
Special tokens file saved in model_output_deeppavlov_rubert/rubist_second_traine
```

```
d/checkpoint-150/special_tokens_map.json
Deleting older checkpoint
[model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-120] due to
args.save_total_limit
Saving model checkpoint to
model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-180
Configuration saved in
model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-180/config.json
Model weights saved in model_output_deeppavlov_rubert/rubist_second_trained/chec
kpoint-180/model.safetensors
tokenizer config file saved in model_output_deeppavlov_rubert/rubist_second_trai
ned/checkpoint-180/tokenizer_config.json
Special tokens file saved in model_output_deeppavlov_rubert/rubist_second_traine
d/checkpoint-180/special_tokens_map.json
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 468
  Batch size = 64
Saving model checkpoint to
model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-180
Configuration saved in
model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-180/config.json
Model weights saved in model_output_deeppavlov_rubert/rubist_second_trained/chec
kpoint-180/model.safetensors
tokenizer config file saved in model_output_deeppavlov_rubert/rubist_second_trai
ned/checkpoint-180/tokenizer_config.json
Special tokens file saved in model_output_deeppavlov_rubert/rubist_second_traine
d/checkpoint-180/special_tokens_map.json


Training completed. Do not forget to share your model on huggingface.co/models
=)


Loading best model from
model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-150 (score:
0.47871941328048706).
Deleting older checkpoint
[model_output_deeppavlov_rubert/rubist_second_trained/checkpoint-180] due to
args.save_total_limit
Saving model checkpoint to model_output_deeppavlov_rubert/rubist_second_trained
Configuration saved in
model_output_deeppavlov_rubert/rubist_second_trained/config.json
```

```
Model weights saved in
model_output_deeppavlov_rubert/rubist_second_trained/model.safetensors
tokenizer config file saved in
model_output_deeppavlov_rubert/rubist_second_trained/tokenizer_config.json
Special tokens file saved in
model_output_deeppavlov_rubert/rubist_second_trained/special_tokens_map.json
loading configuration file
model_output_deeppavlov_rubert/rubist_second_trained/config.json
Model config BertConfig {
  "_name_or_path": "model_output_deeppavlov_rubert/rubist_second_trained",
  "architectures": [
    "BertForSequenceClassification"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_past": true,
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "problem_type": "single_label_classification",
  "torch_dtype": "float32",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 119547
}


loading weights file
model_output_deeppavlov_rubert/rubist_second_trained/model.safetensors

Estimated total emissions: 0.0010944574206333383 kg CO2
Number of unique labels: 2
```

```
All model checkpoint weights were used when initializing
BertForSequenceClassification.

All the weights of BertForSequenceClassification were initialized from the model
checkpoint at model_output_deeppavlov_rubert/rubist_second_trained.
If your task is similar to the task the model of the checkpoint was trained on,
you can already use BertForSequenceClassification for predictions without
further training.
loading file vocab.txt
loading file tokenizer.json
loading file added_tokens.json
loading file special_tokens_map.json
loading file tokenizer_config.json

Map:   0%|          | 0/584 [00:00<?, ? examples/s]

Map:   0%|          | 0/584 [00:00<?, ? examples/s]

Sample tokenized input from test: {'stereotype_type': 'profession', 'text':
'                      ', 'category': 0, 'data_name':
'rubist_second', '__index_level_0__': 80, 'input_ids': [101, 5247, 6345, 44670,
15411, 16038, 74989, 102, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0], 'token_type_ids': [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0], 'attention_mask': [1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], 'labels': 0}
```

[12]:

	precision	recall	f1-score	support
0	0.808458	0.835476	0.821745	389.000000
1	0.648352	0.605128	0.625995	195.000000
accuracy	0.758562	0.758562	0.758562	0.758562
macro avg	0.728405	0.720302	0.723870	584.000000
weighted avg	0.754998	0.758562	0.756383	584.000000

[13]:
```python
from BERT_Models_Fine_Tuning_Russian import (data_loader, train_model,␣
 ↪evaluate_model)

gc.collect()
torch.cuda.empty_cache()

# Load and combine relevant datasets
train_data_rubist, test_data_rubist = data_loader(csv_file_path='COMP0173_Data/
 ↪rubist.csv', labelling_criteria='stereotype', dataset_name='rubist',␣
 ↪sample_size=1000000, num_examples=5)
train_data_rubist_second, test_data_rubist_second =␣
 ↪data_loader(csv_file_path='COMP0173_Data/rubist_second.csv',␣
 ↪labelling_criteria='stereotype', dataset_name='rubist_second',␣
 ↪sample_size=1000000, num_examples=5)

# Execute full pipeline for Deepavlov model
```

```
train_model(train_data_rubist, model_path='ai-forever/ruBert-base',␣
 ↪batch_size=64, epoch=6, learning_rate=2e-5,␣
 ↪model_output_base_dir='model_output_ruberta_base',␣
 ↪dataset_name='rubist_trained', seed=42)
evaluate_model(test_data_rubist, model_output_dir='model_output_ruberta_base/
 ↪rubist_trained', result_output_base_dir='result_output_ruberta_base',␣
 ↪dataset_name='rubist_trained', seed=42)

gc.collect()
torch.cuda.empty_cache()

train_model(train_data_rubist_second, model_path='ai-forever/ruBert-base',␣
 ↪batch_size=64, epoch=6, learning_rate=2e-5,␣
 ↪model_output_base_dir='model_output_ruberta_base',␣
 ↪dataset_name='rubist_second_trained', seed=42)
evaluate_model(test_data_rubist_second,␣
 ↪model_output_dir='model_output_ruberta_base/rubist_second_trained',␣
 ↪result_output_base_dir='result_output_ruberta_base',␣
 ↪dataset_name='rubist_second_trained', seed=42)
```

```
First few examples from the training data:
     stereotype_type                                                text  \
168       profession
2883     nationality                                …
2919           lgbtq                                …
3412      profession                                …
732            lgbtq

      category data_name
168          1    rubist
2883         0    rubist
2919         1    rubist
3412         0    rubist
732          1    rubist
First few examples from the testing data:
     stereotype_type                                                text  \
2726      profession                                …
1066          gender                                …
2547     nationality                                …
1834      profession                                …
1825     nationality                                …

      category data_name
2726         0    rubist
1066         0    rubist
2547         0    rubist
1834         0    rubist
```

```
1825         0    rubist
Train data size: 3372
Test data size: 844
First few examples from the training data:
     stereotype_type                                          text  category  \
1005          gender                                            1
1001          gender                                          0
2619     nationality                                           0
2213          gender                                     0
934           gender                                      0

          data_name
1005  rubist_second
1001  rubist_second
2619  rubist_second
2213  rubist_second
934   rubist_second
First few examples from the testing data:
     stereotype_type                                              text  \
80        profession
1814          gender
2277      profession                                …
61        profession                                …
755           gender                               …

      category      data_name
80           0  rubist_second
1814         0  rubist_second
2277         0  rubist_second
61           0  rubist_second
755          0  rubist_second
Train data size: 2336
Test data size: 584
Number of unique labels: 2

config.json:   0%|          | 0.00/590 [00:00<?, ?B/s]

loading configuration file config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--ai-forever--ruBert-
base/snapshots/05f37a2ca9e333fd18f30cd0c96c68d274793c69/config.json
Model config BertConfig {
  "_name_or_path": "ai-forever/ruBert-base",
  "architectures": [
    "BertForMaskedLM"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
```

```
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 120138
}


pytorch_model.bin:   0%|          | 0.00/716M [00:00<?, ?B/s]

loading weights file pytorch_model.bin from cache at
/home/ec2-user/.cache/huggingface/hub/models--ai-forever--ruBert-
base/snapshots/05f37a2ca9e333fd18f30cd0c96c68d274793c69/pytorch_model.bin
Attempting to create safetensors variant
Attempting to convert .bin model on the fly to safetensors.

model.safetensors:   0%|          | 0.00/716M [00:00<?, ?B/s]

Some weights of the model checkpoint at ai-forever/ruBert-base were not used
when initializing BertForSequenceClassification: ['cls.predictions.bias',
'cls.predictions.decoder.bias', 'cls.predictions.decoder.weight',
'cls.predictions.transform.LayerNorm.bias',
'cls.predictions.transform.LayerNorm.weight',
'cls.predictions.transform.dense.bias',
'cls.predictions.transform.dense.weight', 'cls.seq_relationship.bias',
'cls.seq_relationship.weight']
- This IS expected if you are initializing BertForSequenceClassification from
the checkpoint of a model trained on another task or with another architecture
(e.g. initializing a BertForSequenceClassification model from a
BertForPreTraining model).
- This IS NOT expected if you are initializing BertForSequenceClassification
from the checkpoint of a model that you expect to be exactly identical
(initializing a BertForSequenceClassification model from a
BertForSequenceClassification model).
Some weights of BertForSequenceClassification were not initialized from the
```

```
model checkpoint at ai-forever/ruBert-base and are newly initialized:
['classifier.bias', 'classifier.weight']
You should probably TRAIN this model on a down-stream task to be able to use it
for predictions and inference.
Could not locate the tokenizer configuration file, will try to use the model
config instead.
loading configuration file config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--ai-forever--ruBert-
base/snapshots/05f37a2ca9e333fd18f30cd0c96c68d274793c69/config.json
Model config BertConfig {
  "_name_or_path": "ai-forever/ruBert-base",
  "architectures": [
    "BertForMaskedLM"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 120138
}


vocab.txt: 0.00B [00:00, ?B/s]

loading file vocab.txt from cache at
/home/ec2-user/.cache/huggingface/hub/models--ai-forever--ruBert-
base/snapshots/05f37a2ca9e333fd18f30cd0c96c68d274793c69/vocab.txt
loading file tokenizer.json from cache at None
loading file added_tokens.json from cache at None
loading file special_tokens_map.json from cache at None
```

```
loading file tokenizer_config.json from cache at None
loading configuration file config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--ai-forever--ruBert-
base/snapshots/05f37a2ca9e333fd18f30cd0c96c68d274793c69/config.json
Model config BertConfig {
  "_name_or_path": "ai-forever/ruBert-base",
  "architectures": [
    "BertForMaskedLM"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 120138
}

loading configuration file config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--ai-forever--ruBert-
base/snapshots/05f37a2ca9e333fd18f30cd0c96c68d274793c69/config.json
Model config BertConfig {
  "_name_or_path": "ai-forever/ruBert-base",
  "architectures": [
    "BertForMaskedLM"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
```

```
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 120138
}
```


```
Map:   0%|          | 0/2697 [00:00<?, ? examples/s]
```

```
Map:   0%|          | 0/2697 [00:00<?, ? examples/s]
```


```
Sample tokenized input from train: {'stereotype_type': 'gender', 'text':
'                                        ',
'category': 0, 'data_name': 'rubist', '__index_level_0__': 1317, 'input_ids':
[101, 4269, 12924, 15663, 3822, 16375, 52066, 699, 110, 102821, 5897, 921,
82766, 1306, 102, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], 'token_type_ids': [0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0], 'attention_mask': [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0], 'labels': 0}
```

```
Map:   0%|          | 0/675 [00:00<?, ? examples/s]
```

```
Map:   0%|          | 0/675 [00:00<?, ? examples/s]
```

```
PyTorch: setting up devices
The default value for the training argument `--report_to` will change in v5
(from all installed integrations to none). In v5, you will need to use
`--report_to all` to get the same behavior as now. You should start updating
your code and make this info disappear :-).
```


```
Sample tokenized input from validation: {'stereotype_type': 'gender', 'text':
'                                        ',
'category': 0, 'data_name': 'rubist', '__index_level_0__': 1317, 'input_ids':
[101, 4269, 12924, 15663, 3822, 16375, 52066, 699, 110, 102821, 5897, 921,
82766, 1306, 102, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], 'token_type_ids': [0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
```

```
0], 'attention_mask': [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0]}

The following columns in the training set don't have a corresponding argument in
`BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.
***** Running training *****
  Num examples = 2,697
  Num Epochs = 6
  Instantaneous batch size per device = 64
  Total train batch size (w. parallel, distributed & accumulation) = 64
  Gradient Accumulation steps = 1
  Total optimization steps = 258
  Number of trainable parameters = 178,308,866

<IPython.core.display.HTML object>

The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 675
  Batch size = 64
Saving model checkpoint to
model_output_ruberta_base/rubist_trained/checkpoint-43
Configuration saved in
model_output_ruberta_base/rubist_trained/checkpoint-43/config.json
Model weights saved in
model_output_ruberta_base/rubist_trained/checkpoint-43/model.safetensors
tokenizer config file saved in
model_output_ruberta_base/rubist_trained/checkpoint-43/tokenizer_config.json
Special tokens file saved in
model_output_ruberta_base/rubist_trained/checkpoint-43/special_tokens_map.json
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 675
  Batch size = 64
Saving model checkpoint to
model_output_ruberta_base/rubist_trained/checkpoint-86
```

```
Configuration saved in
model_output_ruberta_base/rubist_trained/checkpoint-86/config.json
Model weights saved in
model_output_ruberta_base/rubist_trained/checkpoint-86/model.safetensors
tokenizer config file saved in
model_output_ruberta_base/rubist_trained/checkpoint-86/tokenizer_config.json
Special tokens file saved in
model_output_ruberta_base/rubist_trained/checkpoint-86/special_tokens_map.json
Deleting older checkpoint
[model_output_ruberta_base/rubist_trained/checkpoint-43] due to
args.save_total_limit
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 675
  Batch size = 64
Saving model checkpoint to
model_output_ruberta_base/rubist_trained/checkpoint-129
Configuration saved in
model_output_ruberta_base/rubist_trained/checkpoint-129/config.json
Model weights saved in
model_output_ruberta_base/rubist_trained/checkpoint-129/model.safetensors
tokenizer config file saved in
model_output_ruberta_base/rubist_trained/checkpoint-129/tokenizer_config.json
Special tokens file saved in
model_output_ruberta_base/rubist_trained/checkpoint-129/special_tokens_map.json
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 675
  Batch size = 64
Saving model checkpoint to
model_output_ruberta_base/rubist_trained/checkpoint-172
Configuration saved in
model_output_ruberta_base/rubist_trained/checkpoint-172/config.json
Model weights saved in
model_output_ruberta_base/rubist_trained/checkpoint-172/model.safetensors
tokenizer config file saved in
model_output_ruberta_base/rubist_trained/checkpoint-172/tokenizer_config.json
Special tokens file saved in
```

```
model_output_ruberta_base/rubist_trained/checkpoint-172/special_tokens_map.json
Deleting older checkpoint
[model_output_ruberta_base/rubist_trained/checkpoint-129] due to
args.save_total_limit
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 675
  Batch size = 64
Saving model checkpoint to
model_output_ruberta_base/rubist_trained/checkpoint-215
Configuration saved in
model_output_ruberta_base/rubist_trained/checkpoint-215/config.json
Model weights saved in
model_output_ruberta_base/rubist_trained/checkpoint-215/model.safetensors
tokenizer config file saved in
model_output_ruberta_base/rubist_trained/checkpoint-215/tokenizer_config.json
Special tokens file saved in
model_output_ruberta_base/rubist_trained/checkpoint-215/special_tokens_map.json
Deleting older checkpoint
[model_output_ruberta_base/rubist_trained/checkpoint-172] due to
args.save_total_limit
Saving model checkpoint to
model_output_ruberta_base/rubist_trained/checkpoint-258
Configuration saved in
model_output_ruberta_base/rubist_trained/checkpoint-258/config.json
Model weights saved in
model_output_ruberta_base/rubist_trained/checkpoint-258/model.safetensors
tokenizer config file saved in
model_output_ruberta_base/rubist_trained/checkpoint-258/tokenizer_config.json
Special tokens file saved in
model_output_ruberta_base/rubist_trained/checkpoint-258/special_tokens_map.json
Deleting older checkpoint
[model_output_ruberta_base/rubist_trained/checkpoint-215] due to
args.save_total_limit
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 675
  Batch size = 64
```

```
Saving model checkpoint to
model_output_ruberta_base/rubist_trained/checkpoint-258
Configuration saved in
model_output_ruberta_base/rubist_trained/checkpoint-258/config.json
Model weights saved in
model_output_ruberta_base/rubist_trained/checkpoint-258/model.safetensors
tokenizer config file saved in
model_output_ruberta_base/rubist_trained/checkpoint-258/tokenizer_config.json
Special tokens file saved in
model_output_ruberta_base/rubist_trained/checkpoint-258/special_tokens_map.json


Training completed. Do not forget to share your model on huggingface.co/models
=)


Loading best model from model_output_ruberta_base/rubist_trained/checkpoint-86
(score: 0.06719378381967545).
Deleting older checkpoint
[model_output_ruberta_base/rubist_trained/checkpoint-258] due to
args.save_total_limit
Saving model checkpoint to model_output_ruberta_base/rubist_trained
Configuration saved in model_output_ruberta_base/rubist_trained/config.json
Model weights saved in
model_output_ruberta_base/rubist_trained/model.safetensors
tokenizer config file saved in
model_output_ruberta_base/rubist_trained/tokenizer_config.json
Special tokens file saved in
model_output_ruberta_base/rubist_trained/special_tokens_map.json
loading configuration file model_output_ruberta_base/rubist_trained/config.json
Model config BertConfig {
  "_name_or_path": "model_output_ruberta_base/rubist_trained",
  "architectures": [
    "BertForSequenceClassification"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
```

```
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "problem_type": "single_label_classification",
  "torch_dtype": "float32",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 120138
}
```


```
loading weights file model_output_ruberta_base/rubist_trained/model.safetensors

Estimated total emissions: 0.0013583508792769448 kg CO2
Number of unique labels: 2

All model checkpoint weights were used when initializing
BertForSequenceClassification.

All the weights of BertForSequenceClassification were initialized from the model
checkpoint at model_output_ruberta_base/rubist_trained.
If your task is similar to the task the model of the checkpoint was trained on,
you can already use BertForSequenceClassification for predictions without
further training.
loading file vocab.txt
loading file tokenizer.json
loading file added_tokens.json
loading file special_tokens_map.json
loading file tokenizer_config.json

Map:   0%|          | 0/844 [00:00<?, ? examples/s]

Map:   0%|          | 0/844 [00:00<?, ? examples/s]

Sample tokenized input from test: {'stereotype_type': 'profession', 'text':
'                                  ', 'category': 0,
'data_name': 'rubist', '__index_level_0__': 2726, 'input_ids': [101, 42890,
1950, 47937, 10088, 815, 16465, 107, 84625, 18777, 734, 74394, 102, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0], 'token_type_ids': [0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], 'attention_mask': [1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
'labels': 0}
Number of unique labels: 2

loading configuration file config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--ai-forever--ruBert-
```

```
base/snapshots/05f37a2ca9e333fd18f30cd0c96c68d274793c69/config.json
Model config BertConfig {
  "_name_or_path": "ai-forever/ruBert-base",
  "architectures": [
    "BertForMaskedLM"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 120138
}

loading weights file pytorch_model.bin from cache at
/home/ec2-user/.cache/huggingface/hub/models--ai-forever--ruBert-
base/snapshots/05f37a2ca9e333fd18f30cd0c96c68d274793c69/pytorch_model.bin
Attempting to create safetensors variant
Attempting to convert .bin model on the fly to safetensors.
Some weights of the model checkpoint at ai-forever/ruBert-base were not used
when initializing BertForSequenceClassification: ['cls.predictions.bias',
'cls.predictions.decoder.bias', 'cls.predictions.decoder.weight',
'cls.predictions.transform.LayerNorm.bias',
'cls.predictions.transform.LayerNorm.weight',
'cls.predictions.transform.dense.bias',
'cls.predictions.transform.dense.weight', 'cls.seq_relationship.bias',
'cls.seq_relationship.weight']
- This IS expected if you are initializing BertForSequenceClassification from
the checkpoint of a model trained on another task or with another architecture
(e.g. initializing a BertForSequenceClassification model from a
```

```
BertForPreTraining model).
- This IS NOT expected if you are initializing BertForSequenceClassification
from the checkpoint of a model that you expect to be exactly identical
(initializing a BertForSequenceClassification model from a
BertForSequenceClassification model).
Some weights of BertForSequenceClassification were not initialized from the
model checkpoint at ai-forever/ruBert-base and are newly initialized:
['classifier.bias', 'classifier.weight']
You should probably TRAIN this model on a down-stream task to be able to use it
for predictions and inference.
Could not locate the tokenizer configuration file, will try to use the model
config instead.
loading configuration file config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--ai-forever--ruBert-
base/snapshots/05f37a2ca9e333fd18f30cd0c96c68d274793c69/config.json
Model config BertConfig {
  "_name_or_path": "ai-forever/ruBert-base",
  "architectures": [
    "BertForMaskedLM"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 120138
}

loading file vocab.txt from cache at
/home/ec2-user/.cache/huggingface/hub/models--ai-forever--ruBert-
```

```
base/snapshots/05f37a2ca9e333fd18f30cd0c96c68d274793c69/vocab.txt
loading file tokenizer.json from cache at None
loading file added_tokens.json from cache at None
loading file special_tokens_map.json from cache at None
loading file tokenizer_config.json from cache at None
loading configuration file config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--ai-forever--ruBert-
base/snapshots/05f37a2ca9e333fd18f30cd0c96c68d274793c69/config.json
Model config BertConfig {
  "_name_or_path": "ai-forever/ruBert-base",
  "architectures": [
    "BertForMaskedLM"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 120138
}

loading configuration file config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--ai-forever--ruBert-
base/snapshots/05f37a2ca9e333fd18f30cd0c96c68d274793c69/config.json
Model config BertConfig {
  "_name_or_path": "ai-forever/ruBert-base",
  "architectures": [
    "BertForMaskedLM"
  ],
  "attention_probs_dropout_prob": 0.1,
```

```
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 120138
}
```


```
Map:   0%|          | 0/1868 [00:00<?, ? examples/s]
```

```
Map:   0%|          | 0/1868 [00:00<?, ? examples/s]
```


```
Sample tokenized input from train: {'stereotype_type': 'nationality', 'text':
'               ', 'category': 0, 'data_name': 'rubist_second',
'__index_level_0__': 1567, 'input_ids': [101, 780, 34330, 1048, 1179, 94517,
102, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
'token_type_ids': [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0], 'attention_mask': [1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], 'labels': 0}
```


```
Map:   0%|          | 0/468 [00:00<?, ? examples/s]
```

```
Map:   0%|          | 0/468 [00:00<?, ? examples/s]
```

```
PyTorch: setting up devices
The default value for the training argument `--report_to` will change in v5
(from all installed integrations to none). In v5, you will need to use
`--report_to all` to get the same behavior as now. You should start updating
your code and make this info disappear :-).
```


```
Sample tokenized input from validation: {'stereotype_type': 'nationality',
'text': '               ', 'category': 0, 'data_name':
'rubist_second', '__index_level_0__': 1567, 'input_ids': [101, 780, 34330, 1048,
```

```
1179, 94517, 102, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
'token_type_ids': [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0], 'attention_mask': [1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], 'labels': 0}

The following columns in the training set don't have a corresponding argument in
`BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.
***** Running training *****
  Num examples = 1,868
  Num Epochs = 6
  Instantaneous batch size per device = 64
  Total train batch size (w. parallel, distributed & accumulation) = 64
  Gradient Accumulation steps = 1
  Total optimization steps = 180
  Number of trainable parameters = 178,308,866

<IPython.core.display.HTML object>

The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 468
  Batch size = 64
Saving model checkpoint to
model_output_ruberta_base/rubist_second_trained/checkpoint-30
Configuration saved in
model_output_ruberta_base/rubist_second_trained/checkpoint-30/config.json
Model weights saved in
model_output_ruberta_base/rubist_second_trained/checkpoint-30/model.safetensors
tokenizer config file saved in model_output_ruberta_base/rubist_second_trained/c
heckpoint-30/tokenizer_config.json
Special tokens file saved in model_output_ruberta_base/rubist_second_trained/che
ckpoint-30/special_tokens_map.json
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 468
  Batch size = 64
```

```
Saving model checkpoint to
model_output_ruberta_base/rubist_second_trained/checkpoint-60
Configuration saved in
model_output_ruberta_base/rubist_second_trained/checkpoint-60/config.json
Model weights saved in
model_output_ruberta_base/rubist_second_trained/checkpoint-60/model.safetensors
tokenizer config file saved in model_output_ruberta_base/rubist_second_trained/c
heckpoint-60/tokenizer_config.json
Special tokens file saved in model_output_ruberta_base/rubist_second_trained/che
ckpoint-60/special_tokens_map.json
Deleting older checkpoint
[model_output_ruberta_base/rubist_second_trained/checkpoint-30] due to
args.save_total_limit
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 468
  Batch size = 64
Saving model checkpoint to
model_output_ruberta_base/rubist_second_trained/checkpoint-90
Configuration saved in
model_output_ruberta_base/rubist_second_trained/checkpoint-90/config.json
Model weights saved in
model_output_ruberta_base/rubist_second_trained/checkpoint-90/model.safetensors
tokenizer config file saved in model_output_ruberta_base/rubist_second_trained/c
heckpoint-90/tokenizer_config.json
Special tokens file saved in model_output_ruberta_base/rubist_second_trained/che
ckpoint-90/special_tokens_map.json
Deleting older checkpoint
[model_output_ruberta_base/rubist_second_trained/checkpoint-60] due to
args.save_total_limit
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 468
  Batch size = 64
Saving model checkpoint to
model_output_ruberta_base/rubist_second_trained/checkpoint-120
Configuration saved in
model_output_ruberta_base/rubist_second_trained/checkpoint-120/config.json
```

```
Model weights saved in
model_output_ruberta_base/rubist_second_trained/checkpoint-120/model.safetensors
tokenizer config file saved in model_output_ruberta_base/rubist_second_trained/c
heckpoint-120/tokenizer_config.json
Special tokens file saved in model_output_ruberta_base/rubist_second_trained/che
ckpoint-120/special_tokens_map.json
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
  Num examples = 468
  Batch size = 64
Saving model checkpoint to
model_output_ruberta_base/rubist_second_trained/checkpoint-150
Configuration saved in
model_output_ruberta_base/rubist_second_trained/checkpoint-150/config.json
Model weights saved in
model_output_ruberta_base/rubist_second_trained/checkpoint-150/model.safetensors
tokenizer config file saved in model_output_ruberta_base/rubist_second_trained/c
heckpoint-150/tokenizer_config.json
Special tokens file saved in model_output_ruberta_base/rubist_second_trained/che
ckpoint-150/special_tokens_map.json
Deleting older checkpoint
[model_output_ruberta_base/rubist_second_trained/checkpoint-120] due to
args.save_total_limit
Saving model checkpoint to
model_output_ruberta_base/rubist_second_trained/checkpoint-180
Configuration saved in
model_output_ruberta_base/rubist_second_trained/checkpoint-180/config.json
Model weights saved in
model_output_ruberta_base/rubist_second_trained/checkpoint-180/model.safetensors
tokenizer config file saved in model_output_ruberta_base/rubist_second_trained/c
heckpoint-180/tokenizer_config.json
Special tokens file saved in model_output_ruberta_base/rubist_second_trained/che
ckpoint-180/special_tokens_map.json
Deleting older checkpoint
[model_output_ruberta_base/rubist_second_trained/checkpoint-150] due to
args.save_total_limit
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`BertForSequenceClassification.forward`,  you can safely ignore this message.

***** Running Evaluation *****
```

```
  Num examples = 468
  Batch size = 64
Saving model checkpoint to
model_output_ruberta_base/rubist_second_trained/checkpoint-180
Configuration saved in
model_output_ruberta_base/rubist_second_trained/checkpoint-180/config.json
Model weights saved in
model_output_ruberta_base/rubist_second_trained/checkpoint-180/model.safetensors
tokenizer config file saved in model_output_ruberta_base/rubist_second_trained/c
heckpoint-180/tokenizer_config.json
Special tokens file saved in model_output_ruberta_base/rubist_second_trained/che
ckpoint-180/special_tokens_map.json


Training completed. Do not forget to share your model on huggingface.co/models
=)


Loading best model from
model_output_ruberta_base/rubist_second_trained/checkpoint-90 (score:
0.45395180583000183).
Deleting older checkpoint
[model_output_ruberta_base/rubist_second_trained/checkpoint-180] due to
args.save_total_limit
Saving model checkpoint to model_output_ruberta_base/rubist_second_trained
Configuration saved in
model_output_ruberta_base/rubist_second_trained/config.json
Model weights saved in
model_output_ruberta_base/rubist_second_trained/model.safetensors
tokenizer config file saved in
model_output_ruberta_base/rubist_second_trained/tokenizer_config.json
Special tokens file saved in
model_output_ruberta_base/rubist_second_trained/special_tokens_map.json
loading configuration file
model_output_ruberta_base/rubist_second_trained/config.json
Model config BertConfig {
  "_name_or_path": "model_output_ruberta_base/rubist_second_trained",
  "architectures": [
    "BertForSequenceClassification"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "directionality": "bidi",
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
```

```
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "pad_token_id": 0,
  "pooler_fc_size": 768,
  "pooler_num_attention_heads": 12,
  "pooler_num_fc_layers": 3,
  "pooler_size_per_head": 128,
  "pooler_type": "first_token_transform",
  "position_embedding_type": "absolute",
  "problem_type": "single_label_classification",
  "torch_dtype": "float32",
  "transformers_version": "4.46.3",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 120138
}
```


```
loading weights file
model_output_ruberta_base/rubist_second_trained/model.safetensors

Estimated total emissions: 0.0010754149601394114 kg CO2
Number of unique labels: 2

All model checkpoint weights were used when initializing
BertForSequenceClassification.

All the weights of BertForSequenceClassification were initialized from the model
checkpoint at model_output_ruberta_base/rubist_second_trained.
If your task is similar to the task the model of the checkpoint was trained on,
you can already use BertForSequenceClassification for predictions without
further training.
loading file vocab.txt
loading file tokenizer.json
loading file added_tokens.json
loading file special_tokens_map.json
loading file tokenizer_config.json

Map:   0%|          | 0/584 [00:00<?, ? examples/s]

Map:   0%|          | 0/584 [00:00<?, ? examples/s]

Sample tokenized input from test: {'stereotype_type': 'profession', 'text':
'                      ', 'category': 0, 'data_name':
'rubist_second', '__index_level_0__': 80, 'input_ids': [101, 1015, 1202, 61810,
1885, 4821, 22860, 102, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0], 'token_type_ids': [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0], 'attention_mask': [1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0,
```

```
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], 'labels': 0}
```

[13]:

	precision	recall	f1-score	support
0	0.820896	0.848329	0.834387	389.000000
1	0.675824	0.630769	0.652520	195.000000
accuracy	0.775685	0.775685	0.775685	0.775685
macro avg	0.748360	0.739549	0.743453	584.000000
weighted avg	0.772456	0.775685	0.773661	584.000000

[ ]:
```
from BERT_Models_Fine_Tuning_Russian import (data_loader, train_model,␣
 ↪evaluate_model)

gc.collect()
torch.cuda.empty_cache()

# Load and combine relevant datasets
train_data_rubist, test_data_rubist = data_loader(csv_file_path='COMP0173_Data/
 ↪rubist.csv', labelling_criteria='stereotype', dataset_name='rubist',␣
 ↪sample_size=1000000, num_examples=5)
train_data_rubist_second, test_data_rubist_second =␣
 ↪data_loader(csv_file_path='COMP0173_Data/rubist_second.csv',␣
 ↪labelling_criteria='stereotype', dataset_name='rubist_second',␣
 ↪sample_size=1000000, num_examples=5)

# Execute full pipeline for Deepavlov model
train_model(train_data_rubist, model_path='FacebookAI/xlm-roberta-base',␣
 ↪batch_size=64, epoch=6, learning_rate=2e-5,␣
 ↪model_output_base_dir='model_output_xlm_roberta_base',␣
 ↪dataset_name='rubist_trained', seed=42)
evaluate_model(test_data_rubist,␣
 ↪model_output_dir='model_output_xlm_roberta_base/rubist_trained',␣
 ↪result_output_base_dir='result_output_xlm_roberta_base',␣
 ↪dataset_name='rubist_trained', seed=42)

gc.collect()
torch.cuda.empty_cache()

train_model(train_data_rubist_second, model_path='FacebookAI/xlm-roberta-base',␣
 ↪batch_size=64, epoch=6, learning_rate=2e-5,␣
 ↪model_output_base_dir='model_output_xlm_roberta_base',␣
 ↪dataset_name='rubist_second_trained', seed=42)
evaluate_model(test_data_rubist_second,␣
 ↪model_output_dir='model_output_xlm_roberta_base/rubist_second_trained',␣
 ↪result_output_base_dir='result_output_xlm_roberta_base',␣
 ↪dataset_name='rubist_second_trained', seed=42)
```

```
First few examples from the training data:
     stereotype_type                                                text  \
```

```
168       profession
2883     nationality                                …
2919           lgbtq                                …
3412      profession                                …
732            lgbtq

      category data_name
168          1    rubist
2883         0    rubist
2919         1    rubist
3412         0    rubist
732          1    rubist
First few examples from the testing data:
     stereotype_type                                                text  \
2726      profession                                …
1066          gender                                …
2547     nationality                                …
1834      profession                                …
1825     nationality                                …

      category data_name
2726         0    rubist
1066         0    rubist
2547         0    rubist
1834         0    rubist
1825         0    rubist
Train data size: 3372
Test data size: 844
First few examples from the training data:
     stereotype_type                                          text  category  \
1005          gender                                             1
1001          gender                                          0
2619     nationality                                           0
2213          gender                                    0
934           gender                                     0

          data_name
1005  rubist_second
1001  rubist_second
2619  rubist_second
2213  rubist_second
934   rubist_second
First few examples from the testing data:
     stereotype_type                                                text  \
80        profession
1814          gender
2277      profession                                …
61        profession                                …
```

```
755           gender                                …

      category      data_name
80           0  rubist_second
1814         0  rubist_second
2277         0  rubist_second
61           0  rubist_second
755          0  rubist_second
Train data size: 2336
Test data size: 584
Number of unique labels: 2
```

```
config.json:   0%|          | 0.00/615 [00:00<?, ?B/s]
```

```
loading configuration file config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--FacebookAI--xlm-roberta-
base/snapshots/e73636d4f797dec63c3081bb6ed5c7b0bb3f2089/config.json
Model config XLMRobertaConfig {
  "_name_or_path": "FacebookAI/xlm-roberta-base",
  "architectures": [
    "XLMRobertaForMaskedLM"
  ],
  "attention_probs_dropout_prob": 0.1,
  "bos_token_id": 0,
  "classifier_dropout": null,
  "eos_token_id": 2,
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-05,
  "max_position_embeddings": 514,
  "model_type": "xlm-roberta",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_past": true,
  "pad_token_id": 1,
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
  "type_vocab_size": 1,
  "use_cache": true,
  "vocab_size": 250002
}
```

```
model.safetensors:   0%|          | 0.00/1.12G [00:00<?, ?B/s]
```

```
loading weights file model.safetensors from cache at
/home/ec2-user/.cache/huggingface/hub/models--FacebookAI--xlm-roberta-
```

```
base/snapshots/e73636d4f797dec63c3081bb6ed5c7b0bb3f2089/model.safetensors
Some weights of the model checkpoint at FacebookAI/xlm-roberta-base were not
used when initializing XLMRobertaForSequenceClassification: ['lm_head.bias',
'lm_head.dense.bias', 'lm_head.dense.weight', 'lm_head.layer_norm.bias',
'lm_head.layer_norm.weight', 'roberta.pooler.dense.bias',
'roberta.pooler.dense.weight']
- This IS expected if you are initializing XLMRobertaForSequenceClassification
from the checkpoint of a model trained on another task or with another
architecture (e.g. initializing a BertForSequenceClassification model from a
BertForPreTraining model).
- This IS NOT expected if you are initializing
XLMRobertaForSequenceClassification from the checkpoint of a model that you
expect to be exactly identical (initializing a BertForSequenceClassification
model from a BertForSequenceClassification model).
Some weights of XLMRobertaForSequenceClassification were not initialized from
the model checkpoint at FacebookAI/xlm-roberta-base and are newly initialized:
['classifier.dense.bias', 'classifier.dense.weight', 'classifier.out_proj.bias',
'classifier.out_proj.weight']
You should probably TRAIN this model on a down-stream task to be able to use it
for predictions and inference.

tokenizer_config.json:   0%|          | 0.00/25.0 [00:00<?, ?B/s]

loading configuration file config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--FacebookAI--xlm-roberta-
base/snapshots/e73636d4f797dec63c3081bb6ed5c7b0bb3f2089/config.json
Model config XLMRobertaConfig {
  "_name_or_path": "FacebookAI/xlm-roberta-base",
  "architectures": [
    "XLMRobertaForMaskedLM"
  ],
  "attention_probs_dropout_prob": 0.1,
  "bos_token_id": 0,
  "classifier_dropout": null,
  "eos_token_id": 2,
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-05,
  "max_position_embeddings": 514,
  "model_type": "xlm-roberta",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_past": true,
  "pad_token_id": 1,
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
```

```
  "type_vocab_size": 1,
  "use_cache": true,
  "vocab_size": 250002
}


sentencepiece.bpe.model:   0%|          | 0.00/5.07M [00:00<?, ?B/s]

tokenizer.json: 0.00B [00:00, ?B/s]

loading file sentencepiece.bpe.model from cache at
/home/ec2-user/.cache/huggingface/hub/models--FacebookAI--xlm-roberta-
base/snapshots/e73636d4f797dec63c3081bb6ed5c7b0bb3f2089/sentencepiece.bpe.model
loading file tokenizer.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--FacebookAI--xlm-roberta-
base/snapshots/e73636d4f797dec63c3081bb6ed5c7b0bb3f2089/tokenizer.json
loading file added_tokens.json from cache at None
loading file special_tokens_map.json from cache at None
loading file tokenizer_config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--FacebookAI--xlm-roberta-
base/snapshots/e73636d4f797dec63c3081bb6ed5c7b0bb3f2089/tokenizer_config.json
loading configuration file config.json from cache at
/home/ec2-user/.cache/huggingface/hub/models--FacebookAI--xlm-roberta-
base/snapshots/e73636d4f797dec63c3081bb6ed5c7b0bb3f2089/config.json
Model config XLMRobertaConfig {
  "_name_or_path": "FacebookAI/xlm-roberta-base",
  "architectures": [
    "XLMRobertaForMaskedLM"
  ],
  "attention_probs_dropout_prob": 0.1,
  "bos_token_id": 0,
  "classifier_dropout": null,
  "eos_token_id": 2,
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-05,
  "max_position_embeddings": 514,
  "model_type": "xlm-roberta",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_past": true,
  "pad_token_id": 1,
  "position_embedding_type": "absolute",
  "transformers_version": "4.46.3",
  "type_vocab_size": 1,
  "use_cache": true,
```

```
  "vocab_size": 250002
}


Map:   0%|          | 0/2697 [00:00<?, ? examples/s]

Map:   0%|          | 0/2697 [00:00<?, ? examples/s]

Sample tokenized input from train: {'stereotype_type': 'gender', 'text':
'                                        ',
'category': 0, 'data_name': 'rubist', '__index_level_0__': 1317, 'input_ids':
[0, 81939, 440, 14276, 4684, 92354, 103, 19816, 2791, 42678, 174783, 4401, 135,
129, 104335, 55533, 86783, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1], 'attention_mask': [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0], 'labels': 0}

Map:   0%|          | 0/675 [00:00<?, ? examples/s]

Map:   0%|          | 0/675 [00:00<?, ? examples/s]

PyTorch: setting up devices

Sample tokenized input from validation: {'stereotype_type': 'gender', 'text':
'                                        ',
'category': 0, 'data_name': 'rubist', '__index_level_0__': 1317, 'input_ids':
[0, 81939, 440, 14276, 4684, 92354, 103, 19816, 2791, 42678, 174783, 4401, 135,
129, 104335, 55533, 86783, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1], 'attention_mask': [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0], 'labels': 0}

The default value for the training argument `--report_to` will change in v5
(from all installed integrations to none). In v5, you will need to use
`--report_to all` to get the same behavior as now. You should start updating
your code and make this info disappear :-).
The following columns in the training set don't have a corresponding argument in
`XLMRobertaForSequenceClassification.forward` and have been ignored: data_name,
__index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`XLMRobertaForSequenceClassification.forward`,  you can safely ignore this
message.
***** Running training *****
  Num examples = 2,697
  Num Epochs = 6
  Instantaneous batch size per device = 64
  Total train batch size (w. parallel, distributed & accumulation) = 64
  Gradient Accumulation steps = 1
  Total optimization steps = 258
  Number of trainable parameters = 278,045,186

<IPython.core.display.HTML object>
```

```
The following columns in the evaluation set don't have a corresponding argument
in `XLMRobertaForSequenceClassification.forward` and have been ignored:
data_name, __index_level_0__, category, text, stereotype_type. If data_name,
__index_level_0__, category, text, stereotype_type are not expected by
`XLMRobertaForSequenceClassification.forward`,  you can safely ignore this
message.

***** Running Evaluation *****
  Num examples = 675
  Batch size = 64
Saving model checkpoint to
model_output_xlm_roberta_base/rubist_trained/checkpoint-43
Configuration saved in
model_output_xlm_roberta_base/rubist_trained/checkpoint-43/config.json
Model weights saved in
model_output_xlm_roberta_base/rubist_trained/checkpoint-43/model.safetensors
tokenizer config file saved in
model_output_xlm_roberta_base/rubist_trained/checkpoint-43/tokenizer_config.json
Special tokens file saved in model_output_xlm_roberta_base/rubist_trained/checkp
oint-43/special_tokens_map.json
```

4 References

[1] Theo King, Zekun Wu, Adriano Koshiyama, Emre Kazim, and Philip Treleaven. 2024. HEARTS: A holistic framework for explainable, sustainable and robust text stereotype detection. arXiv preprint arXiv:2409.11579. Available at: https://arxiv.org/abs/2409.11579 (Accessed: 4 December 2025). https://doi.org/10.48550/arXiv.2409.11579

[2] Theo King, Zekun Wu, Adriano Koshiyama, Emre Kazim, and Philip Treleaven. 2024. HEARTS-Text-Stereotype-Detection (GitHub Repository). Available at: https://github.com/holistic-ai/HEARTS-Text-Stereotype-Detection (Accessed: 4 December 2025).

[3] Theo King, Zekun Wu, Adriano Koshiyama, Emre Kazim, and Philip Treleaven. Holistic AI. 2024. EMGSD: Expanded Multi-Group Stereotype Dataset (HuggingFace Dataset). Available at: https://huggingface.co/datasets/holistic-ai/EMGSD (Accessed: 4 December 2025).

[4] University College London Technical Support Group (TSG). 2025. GPU Access and Usage Documentation. Available at: https://tsg.cs.ucl.ac.uk/gpus/ (Accessed: 6 December 2025).

[5] United Nations. 2025. The 2030 Agenda for Sustainable Development. Available at: https://sdgs.un.org/2030agenda (Accessed: 6 December 2025).

[6] Veronika Grigoreva, Anastasiia Ivanova, Ilseyar Alimova, and Ekaterina Artemova. 2024. RuBia: A Russian Language Bias Detection Dataset. Available at: https://arxiv.org/abs/2403.17553 (Accessed: 9 December 2025).

[7] Veronika Grigoreva, Anastasiia Ivanova, Ilseyar Alimova, and Ekaterina Artemova. 2024. RuBia-Dataset (GitHub Repository). Available at: https://github.com/vergrig/RuBia-Dataset (Accessed: 9 December 2025).

[8] Sismetanin. 2020. Toxic Comments Detection in Russian (GitHub Repository). Available at: https://github.com/sismetanin/toxic-comments-detection-in-russian (Accessed: 9 December 2025).

[9] DeepPavlov. 2019. RuBERT-base-cased (Hugging Face Model). Available at: https://huggingface.co/DeepPavlov/rubert-base-cased (Accessed: 9 December 2025).

[10] AI-Forever. 2023. RuBERT-base (Hugging Face Model). Available at: https://huggingface.co/ai-forever/ruBert-base (Accessed: 9 December 2025).

[11] Hugging Face. 2024. XLM-RoBERTa: Model Documentation. Available at: https://huggingface.co/docs/transformers/en/model_doc/xlm-roberta (Accessed: 9 December 2025).

[12] DeepPavlov. 2020. ruBERT-base-cased-sentence (Hugging Face Model). Available at: https://huggingface.co/DeepPavlov/rubert-base-cased-sentence (Accessed: 9 December 2025).