


```

Downloading https://files.pythonhosted.org/packages/4f/9a/ab96291470e305504aa4b7a2
Collecting ujson
  Downloading https://files.pythonhosted.org/packages/f1/84/e039c6fffc6603f2dfe966972
    |████████████████████| 184kB 48.3MB/s
Collecting boto3
  Downloading https://files.pythonhosted.org/packages/1d/53/c0a4b5dd73c1bf265331e73f
    |████████████████████| 133kB 46.8MB/s
Collecting unicode
  Downloading https://files.pythonhosted.org/packages/d0/42/d9edfed04228bacea2d82490
    |████████████████████| 245kB 50.2MB/s
Collecting ftfy
  Downloading https://files.pythonhosted.org/packages/ff/e2/3b51c53dfffb1e52d9210ebc0
    |████████████████████| 71kB 7.9MB/s
Collecting numpydoc>=0.8.0
  Downloading https://files.pythonhosted.org/packages/60/1d/9e398c53d6ae27d5ab312ddc
    |████████████████████| 51kB 5.3MB/s
Collecting flask-cors>=3.0.7
  Downloading https://files.pythonhosted.org/packages/69/7f/d0aeaaafb5c3c76c8d2141db
Requirement already satisfied: editdistance in /usr/local/lib/python3.6/dist-packages
Collecting word2number>=1.1
  Downloading https://files.pythonhosted.org/packages/4a/29/a31940c848521f0725f0df6b
Requirement already satisfied: scipy in /usr/local/lib/python3.6/dist-packages (from
Collecting responses>=0.7
  Downloading https://files.pythonhosted.org/packages/c1/04/8a5258cfd851c9c89ae5c12c
Collecting conllu==1.3.1
  Downloading https://files.pythonhosted.org/packages/ae/54/b0ae1199f3d01666821b028c
Requirement already satisfied: flask>=1.0.2 in /usr/local/lib/python3.6/dist-package
Collecting parsimonious>=0.8.0
  Downloading https://files.pythonhosted.org/packages/02/5c/067a2f80860a41000a1a7edf

```

```
%cd /content/gen_bert/injecting_numeracy/textual_data_generation
```

```
/content/gen_bert/injecting_numeracy/textual_data_generation
```

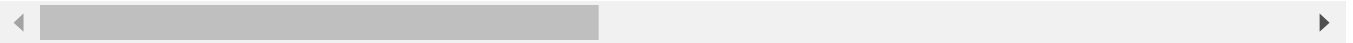
```
!pip install -r requirements.txt
```

```

Requirement already satisfied: numpy in /usr/local/lib/python3.6/dist-packages (from -r
Requirement already satisfied: pandas in /usr/local/lib/python3.6/dist-packages (from -
Requirement already satisfied: spacy==2.1.3 in /usr/local/lib/python3.6/dist-packages (
Requirement already satisfied: tqdm in /usr/local/lib/python3.6/dist-packages (from -r
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.6/dist-packages (
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.6/dist-
Requirement already satisfied: blis<0.3.0,>=0.2.2 in /usr/local/lib/python3.6/dist-pack
Requirement already satisfied: jsonschema<3.0.0,>=2.6.0 in /usr/local/lib/python3.6/dis
Requirement already satisfied: plac<1.0.0,>=0.9.6 in /usr/local/lib/python3.6/dist-pack
Requirement already satisfied: wasabi<1.1.0,>=0.2.0 in /usr/local/lib/python3.6/dist-pa
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.6/dist-pac
Requirement already satisfied: thinc<7.1.0,>=7.0.2 in /usr/local/lib/python3.6/dist-pac
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.6/dist
Requirement already satisfied: srsly<1.1.0,>=0.0.5 in /usr/local/lib/python3.6/dist-pac
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.6/di
Requirement already satisfied: preshed<2.1.0,>=2.0.1 in /usr/local/lib/python3.6/dist-p
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.6/dist-packages (from
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /usr/local/li

```

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.6/dist-pack
 Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.6/dist-packages (
 Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.6/dist-packa



```
!python -m spacy download en
```

```
Collecting en_core_web_sm==2.1.0
  Downloading https://github.com/explosion/spacy-models/releases/download/en\_core\_web\_s
    |████████████████████████████████████████| 11.1MB 3.9MB/s
Building wheels for collected packages: en-core-web-sm
  Building wheel for en-core-web-sm (setup.py) ... done
  Created wheel for en-core-web-sm: filename=en_core_web_sm-2.1.0-cp36-none-any.whl siz
  Stored in directory: /tmp/pip-ephem-wheel-cache-xxwg3sar/wheels/39/ea/3b/507f7df78be8
Successfully built en-core-web-sm
Installing collected packages: en-core-web-sm
  Found existing installation: en-core-web-sm 2.2.5
  Uninstalling en-core-web-sm-2.2.5:
    Successfully uninstalled en-core-web-sm-2.2.5
Successfully installed en-core-web-sm-2.1.0
✓ Download and installation successful
You can now load the model via spacy.load('en_core_web_sm')
✓ Linking successful
/usr/local/lib/python3.6/dist-packages/en_core_web_sm -->
/usr/local/lib/python3.6/dist-packages/spacy/data/en
You can now load the model via spacy.load('en')
```



```
!python generate_examples.py
```

```
100% 16/16 [00:00<00:00, 606.17it/s]
number of generated history questions: 486
cases of empty history examples: 0
number of generated history questions: 0
cases of empty history examples: 0
100% 4/4 [00:00<00:00, 655.54it/s]
number of generated nfl questions: 119
cases of empty nfl examples: 0
number of generated nfl questions: 0
cases of empty nfl examples: 0
total number of generated questions: 605
total cases of empty examples: 0
total number of evaluation questions: 0
```

```
%cd /content/gen_bert/injecting_numeracy/pre_training
```

```
/content/gen_bert/injecting_numeracy/pre_training
```

```
!bash download.sh
```

```
!pwd
```

```
/content/gen_bert/injecting_numeracy/pre_training
```

```
!python gen_train_data_MLM.py --train_corpus ./data/MLM_paras.jsonl --bert_model bert-base-ur
```

```
%cd /content/gen_bert/injecting_numeracy/pre_training/gen_bert
```

```
/content/gen_bert/injecting_numeracy/pre_training/gen_bert
```

```
#drop train feature create
```

```
!python create_examples_n_features.py --split train --drop_json ../data/new_drop_dataset_train
```

```
INFO:pytorch_pretrained_bert.tokenization:loading vocabulary file https://s3.amazonaws.com
INFO:__main__:creating examples
INFO:__main__:Reading file at ../data/new_drop_dataset_train.json
100% 1000/1000 [02:14<00:00, 7.44it/s]
INFO:__main__:Skipped 2 examples, kept 13822 examples.
INFO:__main__:creating features
100% 13822/13822 [01:13<00:00, 188.86it/s]
INFO:__main__:Skipped 0 features, truncated 1019 features, kept 13822 features.
```

```
#drop eval feature create
```

```
!python create_examples_n_features.py --split eval --drop_json ../data/small_dev.json --output
```

```
INFO:pytorch_pretrained_bert.tokenization:loading vocabulary file https://s3.amazonaws.com
INFO:__main__:creating examples
INFO:__main__:Reading file at ../data/small_dev.json
100% 29/29 [00:03<00:00, 8.47it/s]
INFO:__main__:Skipped 0 examples, kept 416 examples.
INFO:__main__:creating features
100% 416/416 [00:02<00:00, 185.81it/s]
INFO:__main__:Skipped 0 features, truncated 0 features, kept 416 features.
```

```
#numeric data train
```

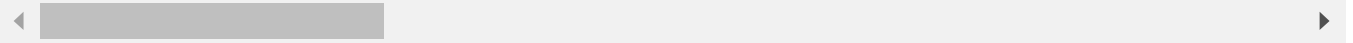
```
!python create_examples_n_features.py --split train --drop_json ../data/synthetic_numeric_train
```

```
INFO:pytorch_pretrained_bert.tokenization:loading vocabulary file https://s3.amazonaws.com
INFO:__main__:creating examples
INFO:__main__:Reading file at ../data/synthetic_numeric_train_drop_format.json
100% 990000/990000 [27:05<00:00, 609.16it/s]
INFO:__main__:Skipped 0 examples, kept 990000 examples.
INFO:__main__:creating features
100% 990000/990000 [05:38<00:00, 2925.50it/s]
INFO:__main__:Skipped 0 features, truncated 5197 features, kept 990000 features.
tcmalloc: large alloc 1073741824 bytes == 0x93402000 @ 0x7f610ecdd1e7 0x5aca7b 0x4e051
```

```
#numeric data eval
```

```
!python create_examples_n_features.py --split eval --drop_json ../data/synthetic_numeric_dev_
```

```
INFO:pytorch_pretrained_bert.tokenization:loading vocabulary file https://s3.amazonaws.com
INFO:__main__:creating examples
INFO:__main__:Reading file at ../data/synthetic_numeric_dev_drop_format.json
100% 9996/9996 [00:16<00:00, 622.39it/s]
INFO:__main__:Skipped 0 examples, kept 9996 examples.
INFO:__main__:creating features
100% 9996/9996 [00:03<00:00, 2908.04it/s]
INFO:__main__:Skipped 0 features, truncated 53 features, kept 9996 features.
```



```
#textual data train
```

```
!python create_examples_n_features.py --split train --drop_json ../data/synthetic_textual_mixed
```

```
#textual data eval
```

```
!python create_examples_n_features.py --split eval --drop_json ../data/synthetic_textual_mixed
```

▼ Pre-Training

```
# GenBERT + ND
```

```
!CUDA_VISIBLE_DEVICES=0,1,2,3,4,5,6,7 python finetune_on_drop.py --do_train --do_eval --n
```

```
/10/2020 01:54:01 - INFO - pytorch_pretrained_bert.tokenization - loading vocabulary f
/10/2020 01:54:01 - INFO - __main__ - device: cpu n_gpu: 0, distributed training: Fals
/10/2020 01:54:02 - INFO - pytorch_pretrained_bert.tokenization - loading vocabulary f
/10/2020 01:54:02 - INFO - modeling - loading weights file https://s3.amazonaws.com/mo
/10/2020 01:54:02 - INFO - modeling - loading configuration file https://s3.amazonaws.com
/10/2020 01:54:02 - INFO - modeling - Model config {
"architectures": [
  "BertForMaskedLM"
],
"attention_probs_dropout_prob": 0.1,
"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,
"type_vocab_size": 2,
"vocab_size": 30522
```

```
/10/2020 01:54:14 - INFO - modeling - Weights of BertTransformer not initialized from
/10/2020 01:54:14 - INFO - modeling - Weights from pretrained model not used in BertTr
/10/2020 01:54:14 - INFO - root - Loading eval examples and features.
```

```

/10/2020 01:54:15 - INFO - __main__ - ***** Running evaluation *****
/10/2020 01:54:15 - INFO - __main__ - Num examples = 9996
/10/2020 01:54:15 - INFO - __main__ - Batch size = 300
/10/2020 01:54:15 - INFO - root - Loading train examples and features.

```

#GenBERT + TD

```
!CUDA_VISIBLE_DEVICES=0,1,2,3,4,5,6,7 python finetune_on_drop.py --do_train --do_eval --
```

#GenBERT + ND + TD

```
!CUDA_VISIBLE_DEVICES=0,1,2,3,4,5,6,7 python train_textual_with_numeric.py --do_train --do
```

▼ Fine-Tuning

#GenBERT + DROP

```
!CUDA_VISIBLE_DEVICES=0,1,2,3 python finetune_on_drop.py --do_train --do_eval --examples
```

```

11/10/2020 03:20:00 - INFO - pytorch_pretrained_bert.tokenization - loading vocabu
11/10/2020 03:20:01 - INFO - __main__ - device: cpu n_gpu: 0, distributed training
11/10/2020 03:20:01 - INFO - pytorch_pretrained_bert.tokenization - loading vocabu
11/10/2020 03:20:01 - INFO - modeling - loading weights file https://s3.amazonaws.com
11/10/2020 03:20:01 - INFO - modeling - loading configuration file https://s3.amazonaws.com
11/10/2020 03:20:01 - INFO - modeling - Model config {
  "architectures": [
    "BertForMaskedLM"
  ],
  "attention_probs_dropout_prob": 0.1,
  "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,
  "type_vocab_size": 2,
  "vocab_size": 30522
}

11/10/2020 03:20:13 - INFO - modeling - Weights of BertTransformer not initialized
11/10/2020 03:20:13 - INFO - modeling - Weights from pretrained model not used in
11/10/2020 03:20:13 - INFO - root - Loading eval examples and features.

11/10/2020 03:20:14 - INFO - __main__ - ***** Running evaluation *****
11/10/2020 03:20:14 - INFO - __main__ - Num examples = 3349

```

```

11/10/2020 03:20:14 - INFO - __main__ - Batch size = 4
11/10/2020 03:20:14 - INFO - root - Loading train examples and features.

11/10/2020 03:20:19 - INFO - __main__ - ***** Running training *****
11/10/2020 03:20:19 - INFO - __main__ - Num examples = 13822
11/10/2020 03:20:19 - INFO - __main__ - Batch size = 4
11/10/2020 03:20:19 - INFO - __main__ - Num steps = 3456
Epoch: 0% 0/1 [00:00<?, ?it/s]
Iteration: 0% 0/3456 [00:00<?, ?it/s]
Iteration: 0% 1/3456 [00:34<33:18:30, 34.71s/it]
Iteration: 0% 2/3456 [01:07<32:47:14, 34.17s/it]
Iteration: 0% 3/3456 [01:39<32:14:34, 33.62s/it]
Iteration: 0% 4/3456 [02:12<31:52:14, 33.24s/it]
Iteration: 0% 5/3456 [02:44<31:36:51, 32.98s/it]
Iteration: 0% 6/3456 [03:17<31:28:05, 32.84s/it]
Iteration: 0% 7/3456 [03:49<31:25:29, 32.80s/it]
Iteration: 0% 8/3456 [04:23<31:32:22, 32.93s/it]
Iteration: 0% 9/3456 [04:55<31:22:11, 32.76s/it]
Iteration: 0% 10/3456 [05:27<31:14:42, 32.64s/it]
Iteration: 0% 11/3456 [06:00<31:05:28, 32.49s/it]
Iteration: 0% 12/3456 [06:32<30:56:29, 32.34s/it]
Iteration: 0% 13/3456 [08:00<47:09:56, 49.32s/it]
Iteration: 0% 14/3456 [09:12<53:28:45, 55.93s/it]
Iteration: 0% 15/3456 [09:44<46:42:40, 48.87s/it]
Iteration: 0% 16/3456 [10:45<38:34:05, 40.36s/it]
Epoch: 0% 0/1 [10:45<?, ?it/s]
Traceback (most recent call last):

```

#GenBERT + ND + DROP

```
!CUDA_VISIBLE_DEVICES=0,1,2,3 python finetune_on_drop.py --do_train --do_eval --examples
```

#GenBERT + TD + DROP

```
!CUDA_VISIBLE_DEVICES=0,1,2,3 python finetune_on_drop.py --do_train --do_eval --examples
```

#GenBERT + ND + TD + DROP

```
!CUDA_VISIBLE_DEVICES=0,1,2,3 python finetune_on_drop.py --do_train --do_eval --examples
```

Inference

```
!CUDA_VISIBLE_DEVICES=0,1,2,3 python finetune_on_drop.py --do_eval --do_inference --examples
```

```

11/10/2020 16:59:19 - INFO - pytorch_pretrained_bert.tokenization - loading vocabular
11/10/2020 16:59:19 - INFO - __main__ - device: cpu n_gpu: 0, distributed training: F
11/10/2020 16:59:19 - INFO - pytorch_pretrained_bert.tokenization - loading vocabular
11/10/2020 16:59:19 - INFO - modeling - loading weights file out_drop_finetune_bert/p
11/10/2020 16:59:19 - INFO - modeling - loading configuration file out_drop_finetune_
11/10/2020 16:59:19 - INFO - modeling - Model config {
    "attention_probs_dropout_prob": 0.1,
    "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,
"num_attention_heads": 12,
"num_hidden_layers": 12,
"type_vocab_size": 2,
"vocab_size": 30522
}

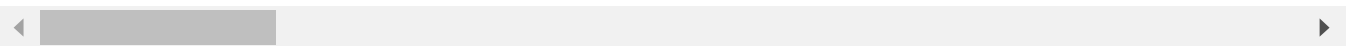
```

11/10/2020 16:59:24 - INFO - root - Loading eval examples and features.

```

11/10/2020 16:59:24 - INFO - __main__ - ***** Running evaluation *****
11/10/2020 16:59:24 - INFO - __main__ - Num examples = 416
11/10/2020 16:59:24 - INFO - __main__ - Batch size = 200
Inference: 0% 0/3 [00:00<?, ?it/s]tcmalloc: large alloc 2516582400 bytes == 0xc2b3600
tcmalloc: large alloc 2516582400 bytes == 0x158b40000 @ 0x7f2710c91b6b 0x7f2710cb1379
tcmalloc: large alloc 2516582400 bytes == 0xc2b36000 @ 0x7f2710c91b6b 0x7f2710cb1379
tcmalloc: large alloc 2516582400 bytes == 0x158b40000 @ 0x7f2710c91b6b 0x7f2710cb1379
tcmalloc: large alloc 2516582400 bytes == 0x158b40000 @ 0x7f2710c91b6b 0x7f2710cb1379
tcmalloc: large alloc 2516582400 bytes == 0xc2b36000 @ 0x7f2710c91b6b 0x7f2710cb1379
tcmalloc: large alloc 2516582400 bytes == 0x158b40000 @ 0x7f2710c91b6b 0x7f2710cb1379
tcmalloc: large alloc 2516582400 bytes == 0xc2b36000 @ 0x7f2710c91b6b 0x7f2710cb1379
Inference: 100% 3/3 [54:40<00:00, 1093.36s/it]
11/10/2020 17:54:04 - INFO - __main__ - EM: 0.3004807692307692, Drop EM: 0.3629807692
11/10/2020 17:54:04 - INFO - __main__ - saving predictions.jsonl in preds
6 7 || 6 1 || 3 5 || 6 3 || 3 3 || 3 6 || Mohnyin nyaungshwe; mone || 3 3 || 3 3 || 17

```



!CUDA_VISIBLE_DEVICES=0,1,2,3 python finetune_on_drop.py --do_eval --do_inference --examples_

!CUDA_VISIBLE_DEVICES=0,1,2,3 python finetune_on_drop.py --do_eval --do_inference --examples_

```

11/10/2020 18:25:34 - INFO - pytorch_pretrained_bert.tokenization - loading vocabular
11/10/2020 18:25:34 - INFO - __main__ - device: cpu n_gpu: 0, distributed training: F
11/10/2020 18:25:35 - INFO - pytorch_pretrained_bert.tokenization - loading vocabular
11/10/2020 18:25:35 - INFO - modeling - loading weights file out_drop_finetune_syntex
11/10/2020 18:25:35 - INFO - modeling - loading configuration file out_drop_finetune_
11/10/2020 18:25:35 - INFO - modeling - Model config {
  "attention_probs_dropout_prob": 0.1,
  "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,
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "type_vocab_size": 2,
  "vocab_size": 30522
}

```

11/10/2020 18:25:47 - INFO - root - Loading eval examples and features.


```

11/10/2020 18:25:47 - INFO - __main__ - ***** Running evaluation *****
11/10/2020 18:25:47 - INFO - __main__ - Num examples = 416
11/10/2020 18:25:47 - INFO - __main__ - Batch size = 200
Inference: 0% 0/3 [00:00<?, ?it/s]tcmalloc: large alloc 2516582400 bytes == 0xc21d800
tcmalloc: large alloc 2516582400 bytes == 0x1581e2000 @ 0x7f5422b4ab6b 0x7f5422b6a379
tcmalloc: large alloc 2516582400 bytes == 0xc21d8000 @ 0x7f5422b4ab6b 0x7f5422b6a379
tcmalloc: large alloc 2516582400 bytes == 0x1581e2000 @ 0x7f5422b4ab6b 0x7f5422b6a379
tcmalloc: large alloc 2516582400 bytes == 0x1581e2000 @ 0x7f5422b4ab6b 0x7f5422b6a379
tcmalloc: large alloc 2516582400 bytes == 0xc21d8000 @ 0x7f5422b4ab6b 0x7f5422b6a379
tcmalloc: large alloc 2516582400 bytes == 0x1581e2000 @ 0x7f5422b4ab6b 0x7f5422b6a379
tcmalloc: large alloc 2516582400 bytes == 0xc21d8000 @ 0x7f5422b4ab6b 0x7f5422b6a379
Inference: 100% 3/3 [54:09<00:00, 1083.28s/it]
11/10/2020 19:19:57 - INFO - __main__ - EM: 0.5360576923076923, Drop EM: 0.6081730769
11/10/2020 19:19:57 - INFO - __main__ - saving predictions.jsonl in preds
6 7 || 1 1 || 5 5 || 3 3 || 3 3 || 6 6 || Shan state of Mone nyaungshwe; mone || 4 3 ||

```

Evaluation

```
#simple bert+drop
```

```
!python drop_eval.py --gold_path ../data/temp_dev.json --prediction_path ./preds/predictions.
```

```

Exact-match accuracy 41.18
F1 score 43.12
41.18 & 43.12
----
date: 1 (5.88%)
  Exact-match accuracy 0.000
  F1 score 33.000
number: 14 (82.35%)
  Exact-match accuracy 42.857
  F1 score 42.857
span: 2 (11.76%)
  Exact-match accuracy 50.000
  F1 score 50.000

```

```
#fintuned with bert + ND + TD
```

```
!python drop_eval.py --gold_path ../data/temp_dev.json --prediction_path ./preds/predictions.
```

```

Exact-match accuracy 64.71
F1 score 66.65
64.71 & 66.65
----
date: 1 (5.88%)
  Exact-match accuracy 0.000
  F1 score 0.000
number: 14 (82.35%)
  Exact-match accuracy 71.429
  F1 score 71.429
span: 2 (11.76%)
  Exact-match accuracy 50.000
  F1 score 66.500

```

