

DMedium_10_3_5_distilbert

August 10, 2022

0.1 Install packages

```
[ ]: !git clone https://github.com/megagonlabs/ditto
      %cd ditto
      !pip install -r requirements.txt
      !git clone https://github.com/rit-git/Snippest_public
      %cd Snippest_public
      !pip install -r requirements.txt
      %cd ..

      import nltk
      nltk.download('stopwords')
```

```
Cloning into 'ditto'...
remote: Enumerating objects: 291, done.
remote: Total 291 (delta 0), reused 0 (delta 0), pack-reused 291
Receiving objects: 100% (291/291), 26.87 MiB | 19.34 MiB/s, done.
Resolving deltas: 100% (142/142), done.
Checking out files: 100% (128/128), done.
/content/ditto
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Collecting gensim==3.8.1
  Downloading gensim-3.8.1-cp37-cp37m-manylinux1_x86_64.whl (24.2 MB)
    |                               | 24.2 MB 90.1 MB/s
Collecting numpy==1.19.2
  Downloading numpy-1.19.2-cp37-cp37m-manylinux2010_x86_64.whl (14.5 MB)
    |                               | 14.5 MB 29.9 MB/s
Collecting regex==2019.12.20
  Downloading regex-2019.12.20-cp37-cp37m-manylinux2010_x86_64.whl (689 kB)
    |                               | 689 kB 69.7 MB/s
Collecting scipy==1.3.2
  Downloading scipy-1.3.2-cp37-cp37m-manylinux1_x86_64.whl (25.2 MB)
    |                               | 25.2 MB 1.3 MB/s
Collecting sentencepiece==0.1.85
  Downloading sentencepiece-0.1.85-cp37-cp37m-manylinux1_x86_64.whl (1.0 MB)
    |                               | 1.0 MB 48.1 MB/s
Requirement already satisfied: sklearn==0.0 in
```

```

/usr/local/lib/python3.7/dist-packages (from -r requirements.txt (line 6)) (0.0)
Collecting spacy==3.1
  Downloading
spacy-3.1.0-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (6.4 MB)
    | 6.4 MB 30.2 MB/s
ERROR: Could not find a version that satisfies the requirement
torch==1.9.0+cu111 (from versions: 1.0.0, 1.0.1, 1.0.1.post2, 1.1.0, 1.2.0,
1.3.0, 1.3.1, 1.4.0, 1.5.0, 1.5.1, 1.6.0, 1.7.0, 1.7.1, 1.8.0, 1.8.1, 1.9.0,
1.9.1, 1.10.0, 1.10.1, 1.10.2, 1.11.0, 1.12.0, 1.12.1)
ERROR: No matching distribution found for torch==1.9.0+cu111
Cloning into 'Snippext_public'...
remote: Enumerating objects: 413, done.
remote: Counting objects: 100% (22/22), done.
remote: Compressing objects: 100% (8/8), done.
remote: Total 413 (delta 15), reused 14 (delta 14), pack-reused 391
Receiving objects: 100% (413/413), 21.06 MiB | 21.43 MiB/s, done.
Resolving deltas: 100% (196/196), done.
/content/ditto/Snippext_public
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Collecting gensim==3.8.1
  Using cached gensim-3.8.1-cp37-cp37m-manylinux1_x86_64.whl (24.2 MB)
Collecting numpy==1.19.2
  Using cached numpy-1.19.2-cp37-cp37m-manylinux2010_x86_64.whl (14.5 MB)
Collecting regex==2019.12.20
  Using cached regex-2019.12.20-cp37-cp37m-manylinux2010_x86_64.whl (689 kB)
Collecting spacy==2.2.3
  Downloading spacy-2.2.3-cp37-cp37m-manylinux1_x86_64.whl (10.4 MB)
    | 10.4 MB 3.5 MB/s
Collecting sentencepiece==0.1.85
  Using cached sentencepiece-0.1.85-cp37-cp37m-manylinux1_x86_64.whl (1.0 MB)
Requirement already satisfied: sklearn==0.0 in /usr/local/lib/python3.7/dist-
packages (from -r requirements.txt (line 6)) (0.0)
Collecting tensorboardX==2.0
  Downloading tensorboardX-2.0-py2.py3-none-any.whl (195 kB)
    | 195 kB 55.5 MB/s
Collecting torch==1.4.0
  Downloading torch-1.4.0-cp37-cp37m-manylinux1_x86_64.whl (753.4 MB)
    | 753.4 MB 6.7 kB/s
Collecting tqdm==4.41.0
  Downloading tqdm-4.41.0-py2.py3-none-any.whl (56 kB)
    | 56 kB 4.2 MB/s
Collecting transformers==3.1.0
  Downloading transformers-3.1.0-py3-none-any.whl (884 kB)
    | 884 kB 68.1 MB/s
Collecting jsonlines==1.2.0

```

```

    Downloading jsonlines-1.2.0-py2.py3-none-any.whl (7.6 kB)
Collecting nltk==3.4.5
    Downloading nltk-3.4.5.zip (1.5 MB)
        | 1.5 MB 50.2 MB/s
Requirement already satisfied: scipy>=0.18.1 in
/usr/local/lib/python3.7/dist-packages (from gensim==3.8.1->-r requirements.txt
(line 1)) (1.7.3)
Requirement already satisfied: six>=1.5.0 in /usr/local/lib/python3.7/dist-
packages (from gensim==3.8.1->-r requirements.txt (line 1)) (1.15.0)
Requirement already satisfied: smart-open>=1.8.1 in
/usr/local/lib/python3.7/dist-packages (from gensim==3.8.1->-r requirements.txt
(line 1)) (5.2.1)
Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in
/usr/local/lib/python3.7/dist-packages (from spacy==2.2.3->-r requirements.txt
(line 4)) (0.10.1)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in
/usr/local/lib/python3.7/dist-packages (from spacy==2.2.3->-r requirements.txt
(line 4)) (3.0.6)
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-
packages (from spacy==2.2.3->-r requirements.txt (line 4)) (57.4.0)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in
/usr/local/lib/python3.7/dist-packages (from spacy==2.2.3->-r requirements.txt
(line 4)) (1.0.7)
Collecting catalogue<1.1.0,>=0.0.7
    Downloading catalogue-1.0.0-py2.py3-none-any.whl (7.7 kB)
Collecting srsly<1.1.0,>=0.1.0
    Downloading srsly-1.0.5-cp37-cp37m-manylinux2014_x86_64.whl (184 kB)
        | 184 kB 68.7 MB/s
Collecting plac<1.2.0,>=0.9.6
    Downloading plac-1.1.3-py2.py3-none-any.whl (20 kB)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in
/usr/local/lib/python3.7/dist-packages (from spacy==2.2.3->-r requirements.txt
(line 4)) (2.0.6)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in
/usr/local/lib/python3.7/dist-packages (from spacy==2.2.3->-r requirements.txt
(line 4)) (2.23.0)
Collecting thinc<7.4.0,>=7.3.0
    Downloading thinc-7.3.1-cp37-cp37m-manylinux1_x86_64.whl (2.2 MB)
        | 2.2 MB 56.4 MB/s
Collecting blis<0.5.0,>=0.4.0
    Downloading blis-0.4.1-cp37-cp37m-manylinux1_x86_64.whl (3.7 MB)
        | 3.7 MB 53.1 MB/s
Requirement already satisfied: scikit-learn in
/usr/local/lib/python3.7/dist-packages (from sklearn==0.0->-r requirements.txt
(line 6)) (1.0.2)
Requirement already satisfied: protobuf>=3.8.0 in /usr/local/lib/python3.7/dist-
packages (from tensorboardX==2.0->-r requirements.txt (line 8)) (3.17.3)
Requirement already satisfied: packaging in /usr/local/lib/python3.7/dist-

```

```

packages (from transformers==3.1.0->-r requirements.txt (line 11)) (21.3)
Collecting tokenizers==0.8.1.rc2
  Downloading tokenizers-0.8.1rc2-cp37-cp37m-manylinux1_x86_64.whl (3.0 MB)
    |                                     | 3.0 MB 50.0 MB/s
Collecting sacremoses
  Downloading sacremoses-0.0.53.tar.gz (880 kB)
    |                                     | 880 kB 57.4 MB/s
Requirement already satisfied: filelock in /usr/local/lib/python3.7/dist-
packages (from transformers==3.1.0->-r requirements.txt (line 11)) (3.7.1)
Requirement already satisfied: importlib-metadata>=0.20 in
/usr/local/lib/python3.7/dist-packages (from
catalogue<1.1.0,>=0.0.7->spacy==2.2.3->-r requirements.txt (line 4)) (4.12.0)
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-
packages (from importlib-
metadata>=0.20->catalogue<1.1.0,>=0.0.7->spacy==2.2.3->-r requirements.txt (line
4)) (3.8.1)
Requirement already satisfied: typing-extensions>=3.6.4 in
/usr/local/lib/python3.7/dist-packages (from importlib-
metadata>=0.20->catalogue<1.1.0,>=0.0.7->spacy==2.2.3->-r requirements.txt (line
4)) (4.1.1)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.7/dist-packages (from
requests<3.0.0,>=2.13.0->spacy==2.2.3->-r requirements.txt (line 4)) (2022.6.15)
Requirement already satisfied: chardet<4,>=3.0.2 in
/usr/local/lib/python3.7/dist-packages (from
requests<3.0.0,>=2.13.0->spacy==2.2.3->-r requirements.txt (line 4)) (3.0.4)
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in
/usr/local/lib/python3.7/dist-packages (from
requests<3.0.0,>=2.13.0->spacy==2.2.3->-r requirements.txt (line 4)) (1.24.3)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-
packages (from requests<3.0.0,>=2.13.0->spacy==2.2.3->-r requirements.txt (line
4)) (2.10)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in
/usr/local/lib/python3.7/dist-packages (from packaging->transformers==3.1.0->-r
requirements.txt (line 11)) (3.0.9)
Requirement already satisfied: click in /usr/local/lib/python3.7/dist-packages
(from sacremoses->transformers==3.1.0->-r requirements.txt (line 11)) (7.1.2)
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages
(from sacremoses->transformers==3.1.0->-r requirements.txt (line 11)) (1.1.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/usr/local/lib/python3.7/dist-packages (from scikit-learn->sklearn==0.0->-r
requirements.txt (line 6)) (3.1.0)
Building wheels for collected packages: nltk, sacremoses
  Building wheel for nltk (setup.py) ... done
  Created wheel for nltk: filename=nltk-3.4.5-py3-none-any.whl size=1449921
sha256=914a81b9ab1b884e3657d5a1e9f8c12c1611b4655646fb85fada2ae3e8e58dcf
  Stored in directory: /root/.cache/pip/wheels/48/8b/7f/473521e0c731c6566d631b28
1f323842bbda9bd819eb9a3ead

```

```

Building wheel for sacremoses (setup.py) ... done
Created wheel for sacremoses: filename=sacremoses-0.0.53-py3-none-any.whl
size=895260
sha256=24db049d2de689e6c98e178f254af537f3887f4d6b2e88a9fd506ebbe4832ea1
Stored in directory: /root/.cache/pip/wheels/87/39/dd/a83eeef36d0bf98e7a4d1933
a4ad2d660295a40613079bafc9
Successfully built nltk sacremoses
Installing collected packages: numpy, tqdm, srsly, regex, plac, blis,
tokenizers, thinc, sentencepiece, sacremoses, catalogue, transformers, torch,
tensorboardX, spacy, nltk, jsonlines, gensim
Attempting uninstall: numpy
  Found existing installation: numpy 1.21.6
  Uninstalling numpy-1.21.6:
    Successfully uninstalled numpy-1.21.6
Attempting uninstall: tqdm
  Found existing installation: tqdm 4.64.0
  Uninstalling tqdm-4.64.0:
    Successfully uninstalled tqdm-4.64.0
Attempting uninstall: srsly
  Found existing installation: srsly 2.4.4
  Uninstalling srsly-2.4.4:
    Successfully uninstalled srsly-2.4.4
Attempting uninstall: regex
  Found existing installation: regex 2022.6.2
  Uninstalling regex-2022.6.2:
    Successfully uninstalled regex-2022.6.2
Attempting uninstall: blis
  Found existing installation: blis 0.7.8
  Uninstalling blis-0.7.8:
    Successfully uninstalled blis-0.7.8
Attempting uninstall: thinc
  Found existing installation: thinc 8.1.0
  Uninstalling thinc-8.1.0:
    Successfully uninstalled thinc-8.1.0
Attempting uninstall: catalogue
  Found existing installation: catalogue 2.0.8
  Uninstalling catalogue-2.0.8:
    Successfully uninstalled catalogue-2.0.8
Attempting uninstall: torch
  Found existing installation: torch 1.12.0+cu113
  Uninstalling torch-1.12.0+cu113:
    Successfully uninstalled torch-1.12.0+cu113
Attempting uninstall: spacy
  Found existing installation: spacy 3.4.1
  Uninstalling spacy-3.4.1:
    Successfully uninstalled spacy-3.4.1
Attempting uninstall: nltk
  Found existing installation: nltk 3.7

```

```

Uninstalling nltk-3.7:
  Successfully uninstalled nltk-3.7
Attempting uninstall: gensim
  Found existing installation: gensim 3.6.0
  Uninstalling gensim-3.6.0:
    Successfully uninstalled gensim-3.6.0
ERROR: pip's dependency resolver does not currently take into account all
the packages that are installed. This behaviour is the source of the following
dependency conflicts.

xarray-einstats 0.2.2 requires numpy>=1.21, but you have numpy 1.19.2 which is
incompatible.

torchvision 0.13.0+cu113 requires torch==1.12.0, but you have torch 1.4.0 which
is incompatible.

torchtext 0.13.0 requires torch==1.12.0, but you have torch 1.4.0 which is
incompatible.

torchaudio 0.12.0+cu113 requires torch==1.12.0, but you have torch 1.4.0 which
is incompatible.

tensorflow 2.8.2+zzzcolab20220719082949 requires numpy>=1.20, but you have numpy
1.19.2 which is incompatible.

panel 0.12.1 requires tqdm>=4.48.0, but you have tqdm 4.41.0 which is
incompatible.

fastai 2.7.7 requires torch<1.13,>=1.7, but you have torch 1.4.0 which is
incompatible.

en-core-web-sm 3.4.0 requires spacy<3.5.0,>=3.4.0, but you have spacy 2.2.3
which is incompatible.

cmdstanpy 1.0.4 requires numpy>=1.21, but you have numpy 1.19.2 which is
incompatible.
Successfully installed blis-0.4.1 catalogue-1.0.0 gensim-3.8.1 jsonlines-1.2.0
nltk-3.4.5 numpy-1.19.2 plac-1.1.3 regex-2019.12.20 sacremoses-0.0.53
sentencepiece-0.1.85 spacy-2.2.3 srsly-1.0.5 tensorboardX-2.0 thinc-7.3.1
tokenizers-0.8.1rc2 torch-1.4.0 tqdm-4.41.0 transformers-3.1.0
/content/ditto

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.

```

```
[ ]: True
```

0.2 Install fp16 optimization

```
[ ]: !git clone https://github.com/NVIDIA/apex
      %cd apex
      !pip install -v --no-cache-dir ./
      %cd ..
```

```
Cloning into 'apex'...
remote: Enumerating objects: 10085, done.
remote: Counting objects: 100% (201/201), done.
remote: Compressing objects: 100% (138/138), done.
remote: Total 10085 (delta 98), reused 131 (delta 59), pack-reused 9884
Receiving objects: 100% (10085/10085), 14.95 MiB | 23.70 MiB/s, done.
Resolving deltas: 100% (6902/6902), done.
/content/ditto/apex
Using pip 21.1.3 from /usr/local/lib/python3.7/dist-packages/pip (python 3.7)
Value for scheme.platlib does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.7/dist-packages
sysconfig: /usr/lib/python3.7/site-packages
Value for scheme.purelib does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.7/dist-packages
sysconfig: /usr/lib/python3.7/site-packages
Value for scheme.headers does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/include/python3.7/UNKNOWN
sysconfig: /usr/include/python3.7m/UNKNOWN
Value for scheme.scripts does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/bin
sysconfig: /usr/bin
Value for scheme.data does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local
sysconfig: /usr
Additional context:
user = False
home = None
root = None
prefix = None
Non-user install because site-packages writable
Created temporary directory: /tmp/pip-ephem-wheel-cache-x_wyhih9
Created temporary directory: /tmp/pip-req-tracker-tvq2jwl2
Initialized build tracking at /tmp/pip-req-tracker-tvq2jwl2
Created build tracker: /tmp/pip-req-tracker-tvq2jwl2
Entered build tracker: /tmp/pip-req-tracker-tvq2jwl2
```

```

Created temporary directory: /tmp/pip-install-0p6d3ibl
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Processing /content/ditto/apex
  Created temporary directory: /tmp/pip-req-build-ffufnq2m
  DEPRECATION: A future pip version will change local packages to be built
in-place without first copying to a temporary directory. We recommend you use
--use-feature=in-tree-build to test your packages with this new behavior before
it becomes the default.

  pip 21.3 will remove support for this functionality. You can find discussion
regarding this at https://github.com/pypa/pip/issues/7555.
  Added file:///content/ditto/apex to build tracker '/tmp/pip-req-tracker-
tvq2jwl2'
  Running setup.py (path:/tmp/pip-req-build-ffufnq2m/setup.py) egg_info for
package from file:///content/ditto/apex
  Created temporary directory: /tmp/pip-pip-egg-info-lft1ncil
  Running command python setup.py egg_info

  torch.__version__ = 1.4.0

  running egg_info
  creating /tmp/pip-pip-egg-info-lft1ncil/apex.egg-info
  writing /tmp/pip-pip-egg-info-lft1ncil/apex.egg-info/PKG-INFO
  writing dependency_links to /tmp/pip-pip-egg-info-lft1ncil/apex.egg-
info/dependency_links.txt
  writing top-level names to /tmp/pip-pip-egg-info-lft1ncil/apex.egg-
info/top_level.txt
  writing manifest file '/tmp/pip-pip-egg-info-lft1ncil/apex.egg-
info/SOURCES.txt'
  adding license file 'LICENSE'
  writing manifest file '/tmp/pip-pip-egg-info-lft1ncil/apex.egg-
info/SOURCES.txt'
  Source in /tmp/pip-req-build-ffufnq2m has version 0.1, which satisfies
requirement apex==0.1 from file:///content/ditto/apex
  Removed apex==0.1 from file:///content/ditto/apex from build tracker
'/tmp/pip-req-tracker-tvq2jwl2'
Created temporary directory: /tmp/pip-unpack-pqrjd17m
Building wheels for collected packages: apex
  Created temporary directory: /tmp/pip-wheel-zbcid9q2
  Building wheel for apex (setup.py) ... Destination directory:
/tmp/pip-wheel-zbcid9q2
  Running command /usr/bin/python3 -u -c 'import io, os, sys, setuptools,
tokenize; sys.argv[0] = ''"/tmp/pip-req-build-ffufnq2m/setup.py''';

```



```
__file__='''/tmp/pip-req-build-ffufnq2m/setup.py''';f = getattr(tokenize,
''''open''', open)(__file__) if os.path.exists(__file__) else
io.StringIO('''from setuptools import setup; setup()''');code =
f.read().replace('''\r\n''', '''\n''');f.close();exec(compile(code,
__file__, '''exec'''))' bdist_wheel -d /tmp/pip-wheel-zbcid9q2
```

```
torch.__version__ = 1.4.0
```

```
running bdist_wheel
running build
running build_py
creating build
creating build/lib
creating build/lib/apex
copying apex/__init__.py -> build/lib/apex
copying apex/_autocast_utils.py -> build/lib/apex
creating build/lib/apex/normalization
copying apex/normalization/__init__.py -> build/lib/apex/normalization
copying apex/normalization/fused_layer_norm.py -> build/lib/apex/normalization
creating build/lib/apex/RNN
copying apex/RNN/__init__.py -> build/lib/apex/RNN
copying apex/RNN/cells.py -> build/lib/apex/RNN
copying apex/RNN/RNNBackend.py -> build/lib/apex/RNN
copying apex/RNN/models.py -> build/lib/apex/RNN
creating build/lib/apex/contrib
copying apex/contrib/__init__.py -> build/lib/apex/contrib
creating build/lib/apex/multi_tensor_apply
copying apex/multi_tensor_apply/__init__.py ->
build/lib/apex/multi_tensor_apply
copying apex/multi_tensor_apply/multi_tensor_apply.py ->
build/lib/apex/multi_tensor_apply
creating build/lib/apex/transformer
copying apex/transformer/__init__.py -> build/lib/apex/transformer
copying apex/transformer/enums.py -> build/lib/apex/transformer
copying apex/transformer/microbatches.py -> build/lib/apex/transformer
copying apex/transformer/log_util.py -> build/lib/apex/transformer
copying apex/transformer/utils.py -> build/lib/apex/transformer
copying apex/transformer/parallel_state.py -> build/lib/apex/transformer
creating build/lib/apex/amp
copying apex/amp/__init__.py -> build/lib/apex/amp
copying apex/amp/opt.py -> build/lib/apex/amp
copying apex/amp/rnn_compat.py -> build/lib/apex/amp
copying apex/amp/_process_optimizer.py -> build/lib/apex/amp
copying apex/amp/compat.py -> build/lib/apex/amp
copying apex/amp/__version__.py -> build/lib/apex/amp
copying apex/amp/utils.py -> build/lib/apex/amp
```

```

copying apex/amp/scaler.py -> build/lib/apex/amp
copying apex/amp/wrap.py -> build/lib/apex/amp
copying apex/amp/frontend.py -> build/lib/apex/amp
copying apex/amp/_initialize.py -> build/lib/apex/amp
copying apex/amp/handle.py -> build/lib/apex/amp
copying apex/amp/_amp_state.py -> build/lib/apex/amp
copying apex/amp/amp.py -> build/lib/apex/amp
creating build/lib/apex/optimizers
copying apex/optimizers/__init__.py -> build/lib/apex/optimizers
copying apex/optimizers/fused_sgd.py -> build/lib/apex/optimizers
copying apex/optimizers/fused_mixed_precision_lamb.py ->
build/lib/apex/optimizers
copying apex/optimizers/fused_novograd.py -> build/lib/apex/optimizers
copying apex/optimizers/fused_adagrad.py -> build/lib/apex/optimizers
copying apex/optimizers/fused_adam.py -> build/lib/apex/optimizers
copying apex/optimizers/fused_lamb.py -> build/lib/apex/optimizers
creating build/lib/apex/fp16_utils
copying apex/fp16_utils/__init__.py -> build/lib/apex/fp16_utils
copying apex/fp16_utils/loss_scaler.py -> build/lib/apex/fp16_utils
copying apex/fp16_utils/fp16util.py -> build/lib/apex/fp16_utils
copying apex/fp16_utils/fp16_optimizer.py -> build/lib/apex/fp16_utils
creating build/lib/apex/fused_dense
copying apex/fused_dense/__init__.py -> build/lib/apex/fused_dense
copying apex/fused_dense/fused_dense.py -> build/lib/apex/fused_dense
creating build/lib/apex/parallel
copying apex/parallel/optimized_sync_batchnorm.py -> build/lib/apex/parallel
copying apex/parallel/__init__.py -> build/lib/apex/parallel
copying apex/parallel/optimized_sync_batchnorm_kernel.py ->
build/lib/apex/parallel
copying apex/parallel/sync_batchnorm_kernel.py -> build/lib/apex/parallel
copying apex/parallel/multiproc.py -> build/lib/apex/parallel
copying apex/parallel/distributed.py -> build/lib/apex/parallel
copying apex/parallel/LARC.py -> build/lib/apex/parallel
copying apex/parallel/sync_batchnorm.py -> build/lib/apex/parallel
creating build/lib/apex/mlp
copying apex/mlp/__init__.py -> build/lib/apex/mlp
copying apex/mlp/mlp.py -> build/lib/apex/mlp
creating build/lib/apex/contrib/focal_loss
copying apex/contrib/focal_loss/__init__.py ->
build/lib/apex/contrib/focal_loss
copying apex/contrib/focal_loss/focal_loss.py ->
build/lib/apex/contrib/focal_loss
creating build/lib/apex/contrib/peer_memory
copying apex/contrib/peer_memory/__init__.py ->
build/lib/apex/contrib/peer_memory
copying apex/contrib/peer_memory/peer_halo_exchanger_1d.py ->
build/lib/apex/contrib/peer_memory
copying apex/contrib/peer_memory/peer_memory.py ->

```

```

build/lib/apex/contrib/peer_memory
    copying apex/contrib/peer_memory/peer_halo_exchange_module_tests.py ->
build/lib/apex/contrib/peer_memory
    creating build/lib/apex/contrib/groupbn
    copying apex/contrib/groupbn/__init__.py -> build/lib/apex/contrib/groupbn
    copying apex/contrib/groupbn/batch_norm.py -> build/lib/apex/contrib/groupbn
    creating build/lib/apex/contrib/optimizers
    copying apex/contrib/optimizers/__init__.py ->
build/lib/apex/contrib/optimizers
    copying apex/contrib/optimizers/distributed_fused_adam.py ->
build/lib/apex/contrib/optimizers
    copying apex/contrib/optimizers/fused_sgd.py ->
build/lib/apex/contrib/optimizers
    copying apex/contrib/optimizers/fused_adam.py ->
build/lib/apex/contrib/optimizers
    copying apex/contrib/optimizers/fused_lamb.py ->
build/lib/apex/contrib/optimizers
    copying apex/contrib/optimizers/distributed_fused_lamb.py ->
build/lib/apex/contrib/optimizers
    copying apex/contrib/optimizers/fp16_optimizer.py ->
build/lib/apex/contrib/optimizers
    creating build/lib/apex/contrib/multihead_attn
    copying apex/contrib/multihead_attn/__init__.py ->
build/lib/apex/contrib/multihead_attn
    copying apex/contrib/multihead_attn/self_multihead_attn.py ->
build/lib/apex/contrib/multihead_attn
    copying apex/contrib/multihead_attn/fast_self_multihead_attn_norm_add_func.py
-> build/lib/apex/contrib/multihead_attn
    copying apex/contrib/multihead_attn/fast_self_multihead_attn_func.py ->
build/lib/apex/contrib/multihead_attn
    copying apex/contrib/multihead_attn/encdec_multihead_attn.py ->
build/lib/apex/contrib/multihead_attn
    copying apex/contrib/multihead_attn/fast_encdec_multihead_attn_func.py ->
build/lib/apex/contrib/multihead_attn
    copying apex/contrib/multihead_attn/encdec_multihead_attn_func.py ->
build/lib/apex/contrib/multihead_attn
    copying
apex/contrib/multihead_attn/fast_encdec_multihead_attn_norm_add_func.py ->
build/lib/apex/contrib/multihead_attn
    copying apex/contrib/multihead_attn/mask_softmax_dropout_func.py ->
build/lib/apex/contrib/multihead_attn
    copying apex/contrib/multihead_attn/self_multihead_attn_func.py ->
build/lib/apex/contrib/multihead_attn
    creating build/lib/apex/contrib/bottleneck
    copying apex/contrib/bottleneck/__init__.py ->
build/lib/apex/contrib/bottleneck
    copying apex/contrib/bottleneck/bottleneck_module_test.py ->
build/lib/apex/contrib/bottleneck

```

```

    copying apex/contrib/bottleneck/test.py -> build/lib/apex/contrib/bottleneck
    copying apex/contrib/bottleneck/bottleneck.py ->
build/lib/apex/contrib/bottleneck
    copying apex/contrib/bottleneck/halo_exchangers.py ->
build/lib/apex/contrib/bottleneck
    creating build/lib/apex/contrib/sparsity
    copying apex/contrib/sparsity/__init__.py -> build/lib/apex/contrib/sparsity
    copying apex/contrib/sparsity/permutation_lib.py ->
build/lib/apex/contrib/sparsity
    copying apex/contrib/sparsity/asp.py -> build/lib/apex/contrib/sparsity
    copying apex/contrib/sparsity/sparse_masklib.py ->
build/lib/apex/contrib/sparsity
    creating build/lib/apex/contrib/layer_norm
    copying apex/contrib/layer_norm/__init__.py ->
build/lib/apex/contrib/layer_norm
    copying apex/contrib/layer_norm/layer_norm.py ->
build/lib/apex/contrib/layer_norm
    creating build/lib/apex/contrib/transducer
    copying apex/contrib/transducer/__init__.py ->
build/lib/apex/contrib/transducer
    copying apex/contrib/transducer/transducer.py ->
build/lib/apex/contrib/transducer
    creating build/lib/apex/contrib/index_mul_2d
    copying apex/contrib/index_mul_2d/__init__.py ->
build/lib/apex/contrib/index_mul_2d
    copying apex/contrib/index_mul_2d/index_mul_2d.py ->
build/lib/apex/contrib/index_mul_2d
    creating build/lib/apex/contrib/conv_bias_relu
    copying apex/contrib/conv_bias_relu/__init__.py ->
build/lib/apex/contrib/conv_bias_relu
    copying apex/contrib/conv_bias_relu/conv_bias_relu.py ->
build/lib/apex/contrib/conv_bias_relu
    creating build/lib/apex/contrib/xentropy
    copying apex/contrib/xentropy/__init__.py -> build/lib/apex/contrib/xentropy
    copying apex/contrib/xentropy/softmax_xentropy.py ->
build/lib/apex/contrib/xentropy
    creating build/lib/apex/contrib/clip_grad
    copying apex/contrib/clip_grad/__init__.py -> build/lib/apex/contrib/clip_grad
    copying apex/contrib/clip_grad/clip_grad.py ->
build/lib/apex/contrib/clip_grad
    creating build/lib/apex/contrib/fmha
    copying apex/contrib/fmha/__init__.py -> build/lib/apex/contrib/fmha
    copying apex/contrib/fmha/fmha.py -> build/lib/apex/contrib/fmha
    creating build/lib/apex/contrib/sparsity/permutation_search_kernels
    copying apex/contrib/sparsity/permutation_search_kernels/__init__.py ->
build/lib/apex/contrib/sparsity/permutation_search_kernels
    copying apex/contrib/sparsity/permutation_search_kernels/call_permutation_sear
ch_kernels.py -> build/lib/apex/contrib/sparsity/permutation_search_kernels

```

```

    copying
apex/contrib/sparsity/permutation_search_kernels/permutation_utilities.py ->
build/lib/apex/contrib/sparsity/permutation_search_kernels
    copying apex/contrib/sparsity/permutation_search_kernels/exhaustive_search.py
-> build/lib/apex/contrib/sparsity/permutation_search_kernels
    creating build/lib/apex/transformer/tensor_parallel
    copying apex/transformer/tensor_parallel/__init__.py ->
build/lib/apex/transformer/tensor_parallel
    copying apex/transformer/tensor_parallel/mappings.py ->
build/lib/apex/transformer/tensor_parallel
    copying apex/transformer/tensor_parallel/layers.py ->
build/lib/apex/transformer/tensor_parallel
    copying apex/transformer/tensor_parallel/memory.py ->
build/lib/apex/transformer/tensor_parallel
    copying apex/transformer/tensor_parallel/utils.py ->
build/lib/apex/transformer/tensor_parallel
    copying apex/transformer/tensor_parallel/cross_entropy.py ->
build/lib/apex/transformer/tensor_parallel
    copying apex/transformer/tensor_parallel/random.py ->
build/lib/apex/transformer/tensor_parallel
    copying apex/transformer/tensor_parallel/data.py ->
build/lib/apex/transformer/tensor_parallel
    creating build/lib/apex/transformer/testing
    copying apex/transformer/testing/__init__.py ->
build/lib/apex/transformer/testing
    copying apex/transformer/testing/distributed_test_base.py ->
build/lib/apex/transformer/testing
    copying apex/transformer/testing/standalone_bert.py ->
build/lib/apex/transformer/testing
    copying apex/transformer/testing/arguments.py ->
build/lib/apex/transformer/testing
    copying apex/transformer/testing/standalone_transformer_lm.py ->
build/lib/apex/transformer/testing
    copying apex/transformer/testing/commons.py ->
build/lib/apex/transformer/testing
    copying apex/transformer/testing/global_vars.py ->
build/lib/apex/transformer/testing
    copying apex/transformer/testing/standalone_gpt.py ->
build/lib/apex/transformer/testing
    creating build/lib/apex/transformer/functional
    copying apex/transformer/functional/__init__.py ->
build/lib/apex/transformer/functional
    copying apex/transformer/functional/fused_softmax.py ->
build/lib/apex/transformer/functional
    creating build/lib/apex/transformer/amp
    copying apex/transformer/amp/__init__.py -> build/lib/apex/transformer/amp
    copying apex/transformer/amp/grad_scaler.py -> build/lib/apex/transformer/amp
    creating build/lib/apex/transformer/_data

```

```

    copying apex/transformer/_data/__init__.py -> build/lib/apex/transformer/_data
    copying apex/transformer/_data/_batch_sampler.py ->
build/lib/apex/transformer/_data
    creating build/lib/apex/transformer/pipeline_parallel
    copying apex/transformer/pipeline_parallel/__init__.py ->
build/lib/apex/transformer/pipeline_parallel
    copying apex/transformer/pipeline_parallel/_timers.py ->
build/lib/apex/transformer/pipeline_parallel
    copying apex/transformer/pipeline_parallel/utils.py ->
build/lib/apex/transformer/pipeline_parallel
    copying apex/transformer/pipeline_parallel/p2p_communication.py ->
build/lib/apex/transformer/pipeline_parallel
    creating build/lib/apex/transformer/layers
    copying apex/transformer/layers/__init__.py ->
build/lib/apex/transformer/layers
    copying apex/transformer/layers/layer_norm.py ->
build/lib/apex/transformer/layers
    creating build/lib/apex/transformer/pipeline_parallel/schedules
    copying apex/transformer/pipeline_parallel/schedules/fwd_bwd_pipeline_parallel_with_interleaving.py -> build/lib/apex/transformer/pipeline_parallel/schedules
    copying apex/transformer/pipeline_parallel/schedules/__init__.py ->
build/lib/apex/transformer/pipeline_parallel/schedules
    copying apex/transformer/pipeline_parallel/schedules/fwd_bwd_no_pipeline_parallel.py -> build/lib/apex/transformer/pipeline_parallel/schedules
    copying apex/transformer/pipeline_parallel/schedules/common.py ->
build/lib/apex/transformer/pipeline_parallel/schedules
    copying apex/transformer/pipeline_parallel/schedules/fwd_bwd_pipeline_parallel_without_interleaving.py -> build/lib/apex/transformer/pipeline_parallel/schedules
    creating build/lib/apex/amp/lists
    copying apex/amp/lists/__init__.py -> build/lib/apex/amp/lists
    copying apex/amp/lists/tensor_overrides.py -> build/lib/apex/amp/lists
    copying apex/amp/lists/functional_overrides.py -> build/lib/apex/amp/lists
    copying apex/amp/lists/torch_overrides.py -> build/lib/apex/amp/lists
    installing to build/bdist.linux-x86_64/wheel
    running install
    running install_lib
    creating build/bdist.linux-x86_64
    creating build/bdist.linux-x86_64/wheel
    creating build/bdist.linux-x86_64/wheel/apex
    copying build/lib/apex/__init__.py -> build/bdist.linux-x86_64/wheel/apex
    copying build/lib/apex/_autocast_utils.py ->
build/bdist.linux-x86_64/wheel/apex
    creating build/bdist.linux-x86_64/wheel/apex/normalization
    copying build/lib/apex/normalization/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/normalization
    copying build/lib/apex/normalization/fused_layer_norm.py ->
build/bdist.linux-x86_64/wheel/apex/normalization
    creating build/bdist.linux-x86_64/wheel/apex/RNN

```

```

    copying build/lib/apex/RNN/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/RNN
    copying build/lib/apex/RNN/cells.py -> build/bdist.linux-x86_64/wheel/apex/RNN
    copying build/lib/apex/RNN/RNNBackend.py ->
build/bdist.linux-x86_64/wheel/apex/RNN
    copying build/lib/apex/RNN/models.py ->
build/bdist.linux-x86_64/wheel/apex/RNN
    creating build/bdist.linux-x86_64/wheel/apex/contrib
    creating build/bdist.linux-x86_64/wheel/apex/contrib/focal_loss
    copying build/lib/apex/contrib/focal_loss/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/focal_loss
    copying build/lib/apex/contrib/focal_loss/focal_loss.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/focal_loss
    copying build/lib/apex/contrib/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib
    creating build/bdist.linux-x86_64/wheel/apex/contrib/peer_memory
    copying build/lib/apex/contrib/peer_memory/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/peer_memory
    copying build/lib/apex/contrib/peer_memory/peer_halo_exchanger_1d.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/peer_memory
    copying build/lib/apex/contrib/peer_memory/peer_memory.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/peer_memory
    copying build/lib/apex/contrib/peer_memory/peer_halo_exchange_module_tests.py
-> build/bdist.linux-x86_64/wheel/apex/contrib/peer_memory
    creating build/bdist.linux-x86_64/wheel/apex/contrib/groupbn
    copying build/lib/apex/contrib/groupbn/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/groupbn
    copying build/lib/apex/contrib/groupbn/batch_norm.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/groupbn
    creating build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
    copying build/lib/apex/contrib/optimizers/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
    copying build/lib/apex/contrib/optimizers/distributed_fused_adam.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
    copying build/lib/apex/contrib/optimizers/fused_sgd.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
    copying build/lib/apex/contrib/optimizers/fused_adam.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
    copying build/lib/apex/contrib/optimizers/fused_lamb.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
    copying build/lib/apex/contrib/optimizers/distributed_fused_lamb.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
    copying build/lib/apex/contrib/optimizers/fp16_optimizer.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/optimizers
    creating build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying build/lib/apex/contrib/multihead_attn/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying build/lib/apex/contrib/multihead_attn/self_multihead_attn.py ->

```

```

build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying
build/lib/apex/contrib/multihead_attn/fast_self_multihead_attn_norm_add_func.py
-> build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying build/lib/apex/contrib/multihead_attn/fast_self_multihead_attn_func.py
-> build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying build/lib/apex/contrib/multihead_attn/encdec_multihead_attn.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying
build/lib/apex/contrib/multihead_attn/fast_encdec_multihead_attn_func.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying build/lib/apex/contrib/multihead_attn/encdec_multihead_attn_func.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying build/lib/apex/contrib/multihead_attn/fast_encdec_multihead_attn_norm_
add_func.py -> build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying build/lib/apex/contrib/multihead_attn/mask_softmax_dropout_func.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    copying build/lib/apex/contrib/multihead_attn/self_multihead_attn_func.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/multihead_attn
    creating build/bdist.linux-x86_64/wheel/apex/contrib/bottleneck
    copying build/lib/apex/contrib/bottleneck/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/bottleneck
    copying build/lib/apex/contrib/bottleneck/bottleneck_module_test.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/bottleneck
    copying build/lib/apex/contrib/bottleneck/test.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/bottleneck
    copying build/lib/apex/contrib/bottleneck/bottleneck.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/bottleneck
    copying build/lib/apex/contrib/bottleneck/halo_exchangers.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/bottleneck
    creating build/bdist.linux-x86_64/wheel/apex/contrib/sparsity
    copying build/lib/apex/contrib/sparsity/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity
    copying build/lib/apex/contrib/sparsity/permutation_lib.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity
    copying build/lib/apex/contrib/sparsity/asp.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity
    copying build/lib/apex/contrib/sparsity/sparse_masklib.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity
    creating
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity/permutation_search_kernels
    copying build/lib/apex/contrib/sparsity/permutation_search_kernels/__init__.py
->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity/permutation_search_kernels
    copying build/lib/apex/contrib/sparsity/permutation_search_kernels/call_permut
ation_search_kernels.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity/permutation_search_kernels
    copying build/lib/apex/contrib/sparsity/permutation_search_kernels/permutation

```



```

_utilities.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity/permutation_search_kernels
    copying
build/lib/apex/contrib/sparsity/permutation_search_kernels/exhaustive_search.py
->
build/bdist.linux-x86_64/wheel/apex/contrib/sparsity/permutation_search_kernels
    creating build/bdist.linux-x86_64/wheel/apex/contrib/layer_norm
    copying build/lib/apex/contrib/layer_norm/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/layer_norm
    copying build/lib/apex/contrib/layer_norm/layer_norm.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/layer_norm
    creating build/bdist.linux-x86_64/wheel/apex/contrib/transducer
    copying build/lib/apex/contrib/transducer/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/transducer
    copying build/lib/apex/contrib/transducer/transducer.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/transducer
    creating build/bdist.linux-x86_64/wheel/apex/contrib/index_mul_2d
    copying build/lib/apex/contrib/index_mul_2d/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/index_mul_2d
    copying build/lib/apex/contrib/index_mul_2d/index_mul_2d.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/index_mul_2d
    creating build/bdist.linux-x86_64/wheel/apex/contrib/conv_bias_relu
    copying build/lib/apex/contrib/conv_bias_relu/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/conv_bias_relu
    copying build/lib/apex/contrib/conv_bias_relu/conv_bias_relu.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/conv_bias_relu
    creating build/bdist.linux-x86_64/wheel/apex/contrib/xentropy
    copying build/lib/apex/contrib/xentropy/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/xentropy
    copying build/lib/apex/contrib/xentropy/softmax_xentropy.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/xentropy
    creating build/bdist.linux-x86_64/wheel/apex/contrib/clip_grad
    copying build/lib/apex/contrib/clip_grad/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/clip_grad
    copying build/lib/apex/contrib/clip_grad/clip_grad.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/clip_grad
    creating build/bdist.linux-x86_64/wheel/apex/contrib/fmha
    copying build/lib/apex/contrib/fmha/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/fmha
    copying build/lib/apex/contrib/fmha/fmha.py ->
build/bdist.linux-x86_64/wheel/apex/contrib/fmha
    creating build/bdist.linux-x86_64/wheel/apex/multi_tensor_apply
    copying build/lib/apex/multi_tensor_apply/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/multi_tensor_apply
    copying build/lib/apex/multi_tensor_apply/multi_tensor_apply.py ->
build/bdist.linux-x86_64/wheel/apex/multi_tensor_apply
    creating build/bdist.linux-x86_64/wheel/apex/transformer
    copying build/lib/apex/transformer/__init__.py ->

```

```

build/bdist.linux-x86_64/wheel/apex/transformer
  creating build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/mappings.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/layers.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/memory.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/utils.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/cross_entropy.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/random.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/tensor_parallel/data.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/tensor_parallel
  copying build/lib/apex/transformer/enums.py ->
build/bdist.linux-x86_64/wheel/apex/transformer
  copying build/lib/apex/transformer/microbatches.py ->
build/bdist.linux-x86_64/wheel/apex/transformer
  creating build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/testing/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/testing/distributed_test_base.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/testing/standalone_bert.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/testing/arguments.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/testing/standalone_transformer_lm.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/testing/commons.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/testing/global_vars.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  copying build/lib/apex/transformer/testing/standalone_gpt.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/testing
  creating build/bdist.linux-x86_64/wheel/apex/transformer/functional
  copying build/lib/apex/transformer/functional/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/functional
  copying build/lib/apex/transformer/functional/fused_softmax.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/functional
  copying build/lib/apex/transformer/log_util.py ->
build/bdist.linux-x86_64/wheel/apex/transformer
  creating build/bdist.linux-x86_64/wheel/apex/transformer/amp
  copying build/lib/apex/transformer/amp/__init__.py ->

```

```

build/bdist.linux-x86_64/wheel/apex/transformer/amp
  copying build/lib/apex/transformer/amp/grad_scaler.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/amp
  copying build/lib/apex/transformer/utils.py ->
build/bdist.linux-x86_64/wheel/apex/transformer
  creating build/bdist.linux-x86_64/wheel/apex/transformer/_data
  copying build/lib/apex/transformer/_data/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/_data
  copying build/lib/apex/transformer/_data/_batch_sampler.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/_data
  creating build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel
  copying build/lib/apex/transformer/pipeline_parallel/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel
  copying build/lib/apex/transformer/pipeline_parallel/_timers.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel
  creating
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel/schedules
  copying build/lib/apex/transformer/pipeline_parallel/schedules/fwd_bwd_pipeline_parallel_with_interleaving.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel/schedules
  copying build/lib/apex/transformer/pipeline_parallel/schedules/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel/schedules
  copying
build/lib/apex/transformer/pipeline_parallel/schedules/fwd_bwd_no_pipeline_parallel.py
-> build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel/schedules
  copying build/lib/apex/transformer/pipeline_parallel/schedules/common.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel/schedules
  copying build/lib/apex/transformer/pipeline_parallel/schedules/fwd_bwd_pipeline_parallel_without_interleaving.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel/schedules
  copying build/lib/apex/transformer/pipeline_parallel/utils.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel
  copying build/lib/apex/transformer/pipeline_parallel/p2p_communication.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/pipeline_parallel
  creating build/bdist.linux-x86_64/wheel/apex/transformer/layers
  copying build/lib/apex/transformer/layers/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/layers
  copying build/lib/apex/transformer/layers/layer_norm.py ->
build/bdist.linux-x86_64/wheel/apex/transformer/layers
  copying build/lib/apex/transformer/parallel_state.py ->
build/bdist.linux-x86_64/wheel/apex/transformer
  creating build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/opt.py -> build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/rnn_compat.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/_process_optimizer.py ->

```

```

build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/compat.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  creating build/bdist.linux-x86_64/wheel/apex/amp/lists
  copying build/lib/apex/amp/lists/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/amp/lists
  copying build/lib/apex/amp/lists/tensor_overrides.py ->
build/bdist.linux-x86_64/wheel/apex/amp/lists
  copying build/lib/apex/amp/lists/functional_overrides.py ->
build/bdist.linux-x86_64/wheel/apex/amp/lists
  copying build/lib/apex/amp/lists/torch_overrides.py ->
build/bdist.linux-x86_64/wheel/apex/amp/lists
  copying build/lib/apex/amp/__version__.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/utils.py -> build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/scaler.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/wrap.py -> build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/frontend.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/_initialize.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/handle.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/_amp_state.py ->
build/bdist.linux-x86_64/wheel/apex/amp
  copying build/lib/apex/amp/amp.py -> build/bdist.linux-x86_64/wheel/apex/amp
  creating build/bdist.linux-x86_64/wheel/apex/optimizers
  copying build/lib/apex/optimizers/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/optimizers
  copying build/lib/apex/optimizers/fused_sgd.py ->
build/bdist.linux-x86_64/wheel/apex/optimizers
  copying build/lib/apex/optimizers/fused_mixed_precision_lamb.py ->
build/bdist.linux-x86_64/wheel/apex/optimizers
  copying build/lib/apex/optimizers/fused_novograd.py ->
build/bdist.linux-x86_64/wheel/apex/optimizers
  copying build/lib/apex/optimizers/fused_adagrad.py ->
build/bdist.linux-x86_64/wheel/apex/optimizers
  copying build/lib/apex/optimizers/fused_adam.py ->
build/bdist.linux-x86_64/wheel/apex/optimizers
  copying build/lib/apex/optimizers/fused_lamb.py ->
build/bdist.linux-x86_64/wheel/apex/optimizers
  creating build/bdist.linux-x86_64/wheel/apex/fp16_utils
  copying build/lib/apex/fp16_utils/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/fp16_utils
  copying build/lib/apex/fp16_utils/loss_scaler.py ->
build/bdist.linux-x86_64/wheel/apex/fp16_utils
  copying build/lib/apex/fp16_utils/fp16util.py ->

```

```

build/bdist.linux-x86_64/wheel/apex/fp16_utils
  copying build/lib/apex/fp16_utils/fp16_optimizer.py ->
build/bdist.linux-x86_64/wheel/apex/fp16_utils
  creating build/bdist.linux-x86_64/wheel/apex/fused_dense
  copying build/lib/apex/fused_dense/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/fused_dense
  copying build/lib/apex/fused_dense/fused_dense.py ->
build/bdist.linux-x86_64/wheel/apex/fused_dense
  creating build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/optimized_sync_batchnorm.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/optimized_sync_batchnorm_kernel.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/sync_batchnorm_kernel.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/multiproc.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/distributed.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/LARC.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/sync_batchnorm.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  creating build/bdist.linux-x86_64/wheel/apex/mlp
  copying build/lib/apex/mlp/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/mlp
  copying build/lib/apex/mlp/mlp.py -> build/bdist.linux-x86_64/wheel/apex/mlp
running install_egg_info
running egg_info
creating apex.egg-info
writing apex.egg-info/PKG-INFO
writing dependency_links to apex.egg-info/dependency_links.txt
writing top-level names to apex.egg-info/top_level.txt
writing manifest file 'apex.egg-info/SOURCES.txt'
adding license file 'LICENSE'
writing manifest file 'apex.egg-info/SOURCES.txt'
Copying apex.egg-info to build/bdist.linux-x86_64/wheel/apex-0.1-py3.7.egg-
info
running install_scripts
adding license file "LICENSE" (matched pattern "LICEN[CS]E*")
creating build/bdist.linux-x86_64/wheel/apex-0.1.dist-info/WHEEL
creating '/tmp/pip-wheel-zbcid9q2/apex-0.1-py3-none-any.whl' and adding
'build/bdist.linux-x86_64/wheel' to it
  adding 'apex/__init__.py'
  adding 'apex/_autocast_utils.py'
  adding 'apex/RNN/RNNBackend.py'

```

```

adding 'apex/RNN/__init__.py'
adding 'apex/RNN/cells.py'
adding 'apex/RNN/models.py'
adding 'apex/amp/__init__.py'
adding 'apex/amp/__version__.py'
adding 'apex/amp/_amp_state.py'
adding 'apex/amp/_initialize.py'
adding 'apex/amp/_process_optimizer.py'
adding 'apex/amp/amp.py'
adding 'apex/amp/compat.py'
adding 'apex/amp/frontend.py'
adding 'apex/amp/handle.py'
adding 'apex/amp/opt.py'
adding 'apex/amp/rnn_compat.py'
adding 'apex/amp/scaler.py'
adding 'apex/amp/utils.py'
adding 'apex/amp/wrap.py'
adding 'apex/amp/lists/__init__.py'
adding 'apex/amp/lists/functional_overrides.py'
adding 'apex/amp/lists/tensor_overrides.py'
adding 'apex/amp/lists/torch_overrides.py'
adding 'apex/contrib/__init__.py'
adding 'apex/contrib/bottleneck/__init__.py'
adding 'apex/contrib/bottleneck/bottleneck.py'
adding 'apex/contrib/bottleneck/bottleneck_module_test.py'
adding 'apex/contrib/bottleneck/halo_exchangers.py'
adding 'apex/contrib/bottleneck/test.py'
adding 'apex/contrib/clip_grad/__init__.py'
adding 'apex/contrib/clip_grad/clip_grad.py'
adding 'apex/contrib/conv_bias_relu/__init__.py'
adding 'apex/contrib/conv_bias_relu/conv_bias_relu.py'
adding 'apex/contrib/fmha/__init__.py'
adding 'apex/contrib/fmha/fmha.py'
adding 'apex/contrib/focal_loss/__init__.py'
adding 'apex/contrib/focal_loss/focal_loss.py'
adding 'apex/contrib/groupbn/__init__.py'
adding 'apex/contrib/groupbn/batch_norm.py'
adding 'apex/contrib/index_mul_2d/__init__.py'
adding 'apex/contrib/index_mul_2d/index_mul_2d.py'
adding 'apex/contrib/layer_norm/__init__.py'
adding 'apex/contrib/layer_norm/layer_norm.py'
adding 'apex/contrib/multihead_attn/__init__.py'
adding 'apex/contrib/multihead_attn/encdec_multihead_attn.py'
adding 'apex/contrib/multihead_attn/encdec_multihead_attn_func.py'
adding 'apex/contrib/multihead_attn/fast_encdec_multihead_attn_func.py'
adding
'apex/contrib/multihead_attn/fast_encdec_multihead_attn_norm_add_func.py'
    adding 'apex/contrib/multihead_attn/fast_self_multihead_attn_func.py'

```

```

adding 'apex/contrib/multihead_attn/fast_self_multihead_attn_norm_add_func.py'
adding 'apex/contrib/multihead_attn/mask_softmax_dropout_func.py'
adding 'apex/contrib/multihead_attn/self_multihead_attn.py'
adding 'apex/contrib/multihead_attn/self_multihead_attn_func.py'
adding 'apex/contrib/optimizers/__init__.py'
adding 'apex/contrib/optimizers/distributed_fused_adam.py'
adding 'apex/contrib/optimizers/distributed_fused_lamb.py'
adding 'apex/contrib/optimizers/fp16_optimizer.py'
adding 'apex/contrib/optimizers/fused_adam.py'
adding 'apex/contrib/optimizers/fused_lamb.py'
adding 'apex/contrib/optimizers/fused_sgd.py'
adding 'apex/contrib/peer_memory/__init__.py'
adding 'apex/contrib/peer_memory/peer_halo_exchange_module_tests.py'
adding 'apex/contrib/peer_memory/peer_halo_exchanger_1d.py'
adding 'apex/contrib/peer_memory/peer_memory.py'
adding 'apex/contrib/sparsity/__init__.py'
adding 'apex/contrib/sparsity/asp.py'
adding 'apex/contrib/sparsity/permutation_lib.py'
adding 'apex/contrib/sparsity/sparse_masklib.py'
adding 'apex/contrib/sparsity/permutation_search_kernels/__init__.py'
adding 'apex/contrib/sparsity/permutation_search_kernels/call_permutation_sear
ch_kernels.py'
    adding 'apex/contrib/sparsity/permutation_search_kernels/exhaustive_search.py'
    adding
'apex/contrib/sparsity/permutation_search_kernels/permutation_utilities.py'
    adding 'apex/contrib/transducer/__init__.py'
    adding 'apex/contrib/transducer/transducer.py'
    adding 'apex/contrib/xentropy/__init__.py'
    adding 'apex/contrib/xentropy/softmax_xentropy.py'
    adding 'apex/fp16_utils/__init__.py'
    adding 'apex/fp16_utils/fp16_optimizer.py'
    adding 'apex/fp16_utils/fp16util.py'
    adding 'apex/fp16_utils/loss_scaler.py'
    adding 'apex/fused_dense/__init__.py'
    adding 'apex/fused_dense/fused_dense.py'
    adding 'apex/mlp/__init__.py'
    adding 'apex/mlp/mlp.py'
    adding 'apex/multi_tensor_apply/__init__.py'
    adding 'apex/multi_tensor_apply/multi_tensor_apply.py'
    adding 'apex/normalization/__init__.py'
    adding 'apex/normalization/fused_layer_norm.py'
    adding 'apex/optimizers/__init__.py'
    adding 'apex/optimizers/fused_adagrad.py'
    adding 'apex/optimizers/fused_adam.py'
    adding 'apex/optimizers/fused_lamb.py'
    adding 'apex/optimizers/fused_mixed_precision_lamb.py'
    adding 'apex/optimizers/fused_novograd.py'
    adding 'apex/optimizers/fused_sgd.py'

```

```

adding 'apex/parallel/LARC.py'
adding 'apex/parallel/__init__.py'
adding 'apex/parallel/distributed.py'
adding 'apex/parallel/multiproc.py'
adding 'apex/parallel/optimized_sync_batchnorm.py'
adding 'apex/parallel/optimized_sync_batchnorm_kernel.py'
adding 'apex/parallel/sync_batchnorm.py'
adding 'apex/parallel/sync_batchnorm_kernel.py'
adding 'apex/transformer/__init__.py'
adding 'apex/transformer/enums.py'
adding 'apex/transformer/log_util.py'
adding 'apex/transformer/microbatches.py'
adding 'apex/transformer/parallel_state.py'
adding 'apex/transformer/utils.py'
adding 'apex/transformer/_data/__init__.py'
adding 'apex/transformer/_data/_batch_sampler.py'
adding 'apex/transformer/amp/__init__.py'
adding 'apex/transformer/amp/grad_scaler.py'
adding 'apex/transformer/functional/__init__.py'
adding 'apex/transformer/functional/fused_softmax.py'
adding 'apex/transformer/layers/__init__.py'
adding 'apex/transformer/layers/layer_norm.py'
adding 'apex/transformer/pipeline_parallel/__init__.py'
adding 'apex/transformer/pipeline_parallel/_timers.py'
adding 'apex/transformer/pipeline_parallel/p2p_communication.py'
adding 'apex/transformer/pipeline_parallel/utils.py'
adding 'apex/transformer/pipeline_parallel/schedules/__init__.py'
adding 'apex/transformer/pipeline_parallel/schedules/common.py'
adding 'apex/transformer/pipeline_parallel/schedules/fwd_bwd_no_pipelining.py'
adding 'apex/transformer/pipeline_parallel/schedules/fwd_bwd_pipelining_with_i
nterleaving.py'
adding 'apex/transformer/pipeline_parallel/schedules/fwd_bwd_pipelining_withou
t_interleaving.py'
adding 'apex/transformer/tensor_parallel/__init__.py'
adding 'apex/transformer/tensor_parallel/cross_entropy.py'
adding 'apex/transformer/tensor_parallel/data.py'
adding 'apex/transformer/tensor_parallel/layers.py'
adding 'apex/transformer/tensor_parallel/mappings.py'
adding 'apex/transformer/tensor_parallel/memory.py'
adding 'apex/transformer/tensor_parallel/random.py'
adding 'apex/transformer/tensor_parallel/utils.py'
adding 'apex/transformer/testing/__init__.py'
adding 'apex/transformer/testing/arguments.py'
adding 'apex/transformer/testing/commons.py'
adding 'apex/transformer/testing/distributed_test_base.py'
adding 'apex/transformer/testing/global_vars.py'
adding 'apex/transformer/testing/standalone_bert.py'
adding 'apex/transformer/testing/standalone_gpt.py'

```



```

adding 'apex/transformer/testing/standalone_transformer_lm.py'
adding 'apex-0.1.dist-info/LICENSE'
adding 'apex-0.1.dist-info/METADATA'
adding 'apex-0.1.dist-info/WHEEL'
adding 'apex-0.1.dist-info/top_level.txt'
adding 'apex-0.1.dist-info/RECORD'
removing build/bdist.linux-x86_64/wheel
done
Created wheel for apex: filename=apex-0.1-py3-none-any.whl size=298114
sha256=651c2013ace08a6065381fdf603f7eea4612f91f671532f5132a94260ecae4b7
Stored in directory: /tmp/pip-ephem-wheel-cache-
x_wyhih9/wheels/6b/22/a2/1665526ee3c3061243260fd522525eca31398e04bfa5ad7e1f
Successfully built apex
Installing collected packages: apex
Value for scheme.platlib does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.7/dist-packages
sysconfig: /usr/lib/python3.7/site-packages
Value for scheme.purelib does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.7/dist-packages
sysconfig: /usr/lib/python3.7/site-packages
Value for scheme.headers does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/include/python3.7/apex
sysconfig: /usr/include/python3.7m/apex
Value for scheme.scripts does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/bin
sysconfig: /usr/bin
Value for scheme.data does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local
sysconfig: /usr
Additional context:
user = False
home = None
root = None
prefix = None

Value for scheme.platlib does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.7/dist-packages
sysconfig: /usr/lib/python3.7/site-packages
Value for scheme.purelib does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.7/dist-packages
sysconfig: /usr/lib/python3.7/site-packages

```

```

Value for scheme.headers does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/include/python3.7/UNKNOWN
sysconfig: /usr/include/python3.7m/UNKNOWN
Value for scheme.scripts does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/bin
sysconfig: /usr/bin
Value for scheme.data does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local
sysconfig: /usr
Additional context:
user = False
home = None
root = None
prefix = None
Successfully installed apex-0.1
Removed build tracker: '/tmp/pip-req-tracker-tvq2jwl2'
/content/ditto

```

```

[ ]: # some issue with colab
!pip install --upgrade "urllib3==1.25.4" awscli

```

```

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Collecting urllib3==1.25.4
  Downloading urllib3-1.25.4-py2.py3-none-any.whl (125 kB)
    |                               | 125 kB 4.5 MB/s
Collecting awscli
  Downloading awscli-1.25.46-py3-none-any.whl (3.9 MB)
    |                               | 3.9 MB 35.2 MB/s
Collecting botocore==1.27.46
  Downloading botocore-1.27.46-py3-none-any.whl (9.0 MB)
    |                               | 9.0 MB 46.3 MB/s
Collecting rsa<4.8,>=3.1.2
  Downloading rsa-4.7.2-py3-none-any.whl (34 kB)
Collecting s3transfer<0.7.0,>=0.6.0
  Downloading s3transfer-0.6.0-py3-none-any.whl (79 kB)
    |                               | 79 kB 8.6 MB/s
Collecting colorama<0.4.5,>=0.2.5
  Downloading colorama-0.4.4-py2.py3-none-any.whl (16 kB)
Requirement already satisfied: PyYAML<5.5,>=3.10 in
/usr/local/lib/python3.7/dist-packages (from awscli) (3.13)
Collecting docutils<0.17,>=0.10
  Downloading docutils-0.16-py2.py3-none-any.whl (548 kB)
    |                               | 548 kB 38.7 MB/s
Requirement already satisfied: python-dateutil<3.0.0,>=2.1 in

```

```

/usr/local/lib/python3.7/dist-packages (from botocore==1.27.46->awscli) (2.8.2)
Collecting jmespath<2.0.0,>=0.7.1
  Downloading jmespath-1.0.1-py3-none-any.whl (20 kB)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-
packages (from python-dateutil<3.0.0,>=2.1->botocore==1.27.46->awscli) (1.15.0)
Requirement already satisfied: pyasn1>=0.1.3 in /usr/local/lib/python3.7/dist-
packages (from rsa<4.8,>=3.1.2->awscli) (0.4.8)
Installing collected packages: urllib3, jmespath, botocore, s3transfer, rsa,
docutils, colorama, awscli
  Attempting uninstall: urllib3
    Found existing installation: urllib3 1.24.3
    Uninstalling urllib3-1.24.3:
      Successfully uninstalled urllib3-1.24.3
  Attempting uninstall: rsa
    Found existing installation: rsa 4.9
    Uninstalling rsa-4.9:
      Successfully uninstalled rsa-4.9
  Attempting uninstall: docutils
    Found existing installation: docutils 0.17.1
    Uninstalling docutils-0.17.1:
      Successfully uninstalled docutils-0.17.1
ERROR: pip's dependency resolver does not currently take into account all
the packages that are installed. This behaviour is the source of the following
dependency conflicts.

panel 0.12.1 requires tqdm>=4.48.0, but you have tqdm 4.41.0 which is
incompatible.

fastai 2.7.7 requires torch<1.13,>=1.7, but you have torch 1.4.0 which is
incompatible.

Successfully installed awscli-1.25.46 botocore-1.27.46 colorama-0.4.4
docutils-0.16 jmespath-1.0.1 rsa-4.7.2 s3transfer-0.6.0 urllib3-1.25.4

```

```
[ ]: !pip install torch==1.7.1
```

```

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Collecting torch==1.7.1
  Downloading torch-1.7.1-cp37-cp37m-manylinux1_x86_64.whl (776.8 MB)
    |                                     | 776.8 MB 17 kB/s
Requirement already satisfied: typing-extensions in
/usr/local/lib/python3.7/dist-packages (from torch==1.7.1) (4.1.1)
Requirement already satisfied: numpy in /usr/local/lib/python3.7/dist-packages
(from torch==1.7.1) (1.19.2)
Installing collected packages: torch
  Attempting uninstall: torch
    Found existing installation: torch 1.4.0

```

```

Uninstalling torch-1.4.0:
  Successfully uninstalled torch-1.4.0
ERROR: pip's dependency resolver does not currently take into account all
the packages that are installed. This behaviour is the source of the following
dependency conflicts.

torchvision 0.13.0+cu113 requires torch==1.12.0, but you have torch 1.7.1 which
is incompatible.

torchtext 0.13.0 requires torch==1.12.0, but you have torch 1.7.1 which is
incompatible.

torchaudio 0.12.0+cu113 requires torch==1.12.0, but you have torch 1.7.1 which
is incompatible.
Successfully installed torch-1.7.1

```

0.3 importa config.json

```
[ ]: !git clone https://github.com/pauloh48/IC.git
```

```

Cloning into 'IC'...
remote: Enumerating objects: 982, done.
remote: Counting objects: 100% (52/52), done.
remote: Compressing objects: 100% (48/48), done.
remote: Total 982 (delta 16), reused 7 (delta 0), pack-reused 930
Receiving objects: 100% (982/982), 251.89 MiB | 29.72 MiB/s, done.
Resolving deltas: 100% (600/600), done.
Checking out files: 100% (1273/1273), done.

```

0.4 remove e move

```
[ ]: !ls
```

```

apex          ditto.jpg    LICENSE     requirements.txt  run_all_wdc.py
blocking      ditto_light matcher.py  results_ditto    Snippext_public
configs.json  IC          output      run_all_er_magellan.py train_ditto.py
data          input       README.md   run_all_vary_size.py

```

```
[ ]: !rm configs.json
!mv -n /content/ditto/IC/datasEsErros/DMedium_10_3_5/configs.json /content/
↪ditto/
```

1 DA

1.1 DA STRUCTURED

1.1.1 Beer - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Structured/Beer \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --da attr_del \
  --summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Beer.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/Beer \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

```
Downloading: 100% 442/442 [00:00<00:00, 498kB/s]
Downloading: 100% 232k/232k [00:00<00:00, 1.22MB/s]
Downloading: 100% 268M/268M [00:04<00:00, 54.6MB/s]
step: 0, loss: 0.8629884123802185
epoch 1: dev_f1=0.32499999999999996, f1=0.3050847457627119,
best_f1=0.3050847457627119
step: 0, loss: 0.3710261285305023
epoch 2: dev_f1=0.26666666666666667, f1=0.26666666666666667,
best_f1=0.3050847457627119
step: 0, loss: 0.34215229749679565
epoch 3: dev_f1=0.5, f1=0.0, best_f1=0.0
step: 0, loss: 0.3690588176250458
```

```

epoch 4: dev_f1=0.43076923076923085, f1=0.43137254901960786, best_f1=0.0
step: 0, loss: 0.28880953788757324
epoch 5: dev_f1=0.4761904761904762, f1=0.3243243243243243, best_f1=0.0
step: 0, loss: 0.32789039611816406
epoch 6: dev_f1=0.46153846153846156, f1=0.3333333333333333, best_f1=0.0
step: 0, loss: 0.28279969096183777
epoch 7: dev_f1=0.4615384615384615, f1=0.34285714285714286, best_f1=0.0
step: 0, loss: 0.34962835907936096
epoch 8: dev_f1=0.5, f1=0.1333333333333333, best_f1=0.0
step: 0, loss: 0.2532985806465149
epoch 9: dev_f1=0.6206896551724138, f1=0.23076923076923075,
best_f1=0.23076923076923075
step: 0, loss: 0.2561829388141632
epoch 10: dev_f1=0.5405405405405405, f1=0.4285714285714286,
best_f1=0.23076923076923075
step: 0, loss: 0.2617719769477844
epoch 11: dev_f1=0.5714285714285714, f1=0.4285714285714286,
best_f1=0.23076923076923075
step: 0, loss: 0.26442965865135193
epoch 12: dev_f1=0.5806451612903226, f1=0.5000000000000001,
best_f1=0.23076923076923075
step: 0, loss: 0.1525242030620575
epoch 13: dev_f1=0.588235294117647, f1=0.4736842105263159,
best_f1=0.23076923076923075
step: 0, loss: 0.22704198956489563
epoch 14: dev_f1=0.6206896551724138, f1=0.5294117647058824,
best_f1=0.23076923076923075
step: 0, loss: 0.32108840346336365
epoch 15: dev_f1=0.6206896551724138, f1=0.5294117647058824,
best_f1=0.23076923076923075
mv: cannot stat '*_dev.pt': No such file or directory
Traceback (most recent call last):
  File "matcher.py", line 315, in <module>
    hp.lm, hp.use_gpu, hp.fp16)
  File "matcher.py", line 274, in load_model
    raise ModelNotFoundError(checkpoint)
ditto_light.exceptions.ModelNotFoundError: Model
checkpoints/Structured/Beer/model.pt was not found

```

1.1.2 DBLP-ACM - Running the matcher

```

[ ]: # Step 1:
      # train the wdc model with the save_model flag on
      # it will produce two files *_dev.pt and *_test.pt which are the
      # best checkpoints on the dev set and the test set.
      !CUDA_VISIBLE_DEVICES=0 python train_ditto.py \

```

```

--task Structured/DBLP-ACM \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--da entry_swap \
--da span_shuffle \
--summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-ACM.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/DBLP-ACM \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly

```

because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8160415887832642
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4755105972290039
step: 20, loss: 0.5903710126876831
step: 30, loss: 0.41057834029197693
step: 40, loss: 0.17476174235343933
step: 50, loss: 0.10876838117837906
step: 60, loss: 0.06188451498746872
step: 70, loss: 0.17336854338645935
step: 80, loss: 0.19936345517635345
step: 90, loss: 0.020926333963871002
step: 100, loss: 0.020367244258522987
step: 110, loss: 0.10585590451955795
step: 120, loss: 0.06500322371721268
step: 130, loss: 0.006146510597318411
step: 140, loss: 0.04440737143158913
step: 150, loss: 0.11795559525489807
step: 160, loss: 0.08974465727806091
step: 170, loss: 0.014061160385608673
step: 180, loss: 0.005545958410948515
step: 190, loss: 0.010889827273786068
step: 200, loss: 0.0036885105073451996
step: 210, loss: 0.01616021804511547
step: 220, loss: 0.041628025472164154
step: 230, loss: 0.002775606233626604
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9796380090497738, f1=0.9794520547945206,

```



```

best_f1=0.9794520547945206
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.017061227932572365
step: 10, loss: 0.0008350205607712269
step: 20, loss: 0.0024297910276800394
step: 30, loss: 0.0010069587733596563
step: 40, loss: 0.002117197960615158
step: 50, loss: 0.005487262737005949
step: 60, loss: 0.005602821707725525
step: 70, loss: 0.01404603198170662
step: 80, loss: 0.027598967775702477
step: 90, loss: 0.0020454830955713987
step: 100, loss: 0.050290241837501526
step: 110, loss: 0.002467900514602661
step: 120, loss: 0.0017566451570019126
step: 130, loss: 0.0012160944752395153
step: 140, loss: 0.15815457701683044
step: 150, loss: 0.013445201329886913
step: 160, loss: 0.04256366193294525
step: 170, loss: 0.006300853565335274
step: 180, loss: 0.028103269636631012
step: 190, loss: 0.04753812775015831
step: 200, loss: 0.001692092977464199
step: 210, loss: 0.04672050103545189
step: 220, loss: 0.0006692995666526258
step: 230, loss: 0.00228260294534266
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9864253393665158, f1=0.9774774774774775,
best_f1=0.9774774774774775
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.041916653513908386
step: 10, loss: 0.051347121596336365
step: 20, loss: 0.006770884618163109
step: 30, loss: 0.0037178772035986185
step: 40, loss: 0.0030174062121659517
step: 50, loss: 0.004956803284585476
step: 60, loss: 0.00525439390912652
step: 70, loss: 0.0009295502677559853

```

```

step: 80, loss: 0.0011578435078263283
step: 90, loss: 0.000766099663451314
step: 100, loss: 0.0009835881646722555
step: 110, loss: 0.0010470045963302255
step: 120, loss: 0.01669267751276493
step: 130, loss: 0.001162037719041109
step: 140, loss: 0.0010365961352363229
step: 150, loss: 0.0017913891933858395
step: 160, loss: 0.04562872648239136
step: 170, loss: 0.003807427827268839
step: 180, loss: 0.0022824916522949934
step: 190, loss: 0.0019927460234612226
step: 200, loss: 0.0009835169184952974
step: 210, loss: 0.03457876294851303
step: 220, loss: 0.0008485105354338884
step: 230, loss: 0.05178677290678024
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.9775280898876404, f1=0.9696969696969697,
best_f1=0.9774774774774775
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.035807304084300995
step: 10, loss: 0.0006769067258574069
step: 20, loss: 0.0429256409406662
step: 30, loss: 0.00031616329215466976
step: 40, loss: 0.002403078367933631
step: 50, loss: 0.006635364610701799
step: 60, loss: 0.0011184830218553543
step: 70, loss: 0.03271908685564995
step: 80, loss: 0.09567373991012573
step: 90, loss: 0.01284290011972189
step: 100, loss: 0.0022051967680454254
step: 110, loss: 0.006567707750946283
step: 120, loss: 0.006140292156487703
step: 130, loss: 0.0018258095951750875
step: 140, loss: 0.0011319321347400546
step: 150, loss: 0.0009931178065016866
step: 160, loss: 0.0006034817779436707
step: 170, loss: 0.0157785601913929
step: 180, loss: 0.03809460252523422
step: 190, loss: 0.0005393340834416449
step: 200, loss: 0.0006228232523426414
step: 210, loss: 0.00024982969625853
step: 220, loss: 0.0005608896026387811
step: 230, loss: 0.0006401300779543817

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.9854423292273236, f1=0.9831649831649831,
best_f1=0.9774774774774775
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004957186174578965
step: 10, loss: 0.00036004596040584147
step: 20, loss: 0.0004581224638968706
step: 30, loss: 0.0002065292646875605
step: 40, loss: 0.00048424338456243277
step: 50, loss: 0.0002844925329554826
step: 60, loss: 0.001282182172872126
step: 70, loss: 0.00026802896172739565
step: 80, loss: 0.0004954335745424032
step: 90, loss: 0.0041720508597791195
step: 100, loss: 0.003368562087416649
step: 110, loss: 0.0005451760371215641
step: 120, loss: 0.05584924668073654
step: 130, loss: 0.11399608105421066
step: 140, loss: 0.0015224883100017905
step: 150, loss: 0.00017352694703731686
step: 160, loss: 0.01629571057856083
step: 170, loss: 0.04291504994034767
step: 180, loss: 0.0005509341135621071
step: 190, loss: 0.003082918468862772
step: 200, loss: 0.003329755272716284
step: 210, loss: 0.00035824786755256355
step: 220, loss: 0.0004051959840580821
step: 230, loss: 0.0009051524102687836
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.9887387387387387, f1=0.9841986455981941,
best_f1=0.9841986455981941
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006547418888658285
step: 10, loss: 0.0003733580815605819
step: 20, loss: 0.00043600655044429004
step: 30, loss: 0.0004515799228101969

```

```

step: 40, loss: 0.0011670500971376896
step: 50, loss: 0.002817238215357065
step: 60, loss: 0.01639794558286667
step: 70, loss: 0.00037307324237190187
step: 80, loss: 0.0024902704171836376
step: 90, loss: 0.0005512584466487169
step: 100, loss: 0.0005005231942050159
step: 110, loss: 0.063006691634655
step: 120, loss: 0.00017747087986208498
step: 130, loss: 0.008630816824734211
step: 140, loss: 0.005420682951807976
step: 150, loss: 0.00023613755183760077
step: 160, loss: 0.0005486375303007662
step: 170, loss: 0.0002322673099115491
step: 180, loss: 0.005109469406306744
step: 190, loss: 0.00024122757895383984
step: 200, loss: 0.0002445668214932084
step: 210, loss: 0.0008641376043669879
step: 220, loss: 0.0009127047378569841
step: 230, loss: 0.00018642409122548997
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 6: dev_f1=0.9887640449438202, f1=0.9853438556933484,
best_f1=0.9853438556933484
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03762812539935112
step: 10, loss: 0.00018776647630147636
step: 20, loss: 0.0001370481913909316
step: 30, loss: 0.0002592384989839047
step: 40, loss: 0.0004924170207232237
step: 50, loss: 0.00010438496974529698
step: 60, loss: 0.0003791565541177988
step: 70, loss: 0.002967463806271553
step: 80, loss: 0.00013478592154569924
step: 90, loss: 0.00020949837926309556
step: 100, loss: 0.00020969707111362368
step: 110, loss: 0.0223720520734787
step: 120, loss: 0.0019167461432516575
step: 130, loss: 0.0018843277357518673
step: 140, loss: 0.0006591046112589538
step: 150, loss: 0.0008156835101544857

```

```

step: 160, loss: 0.0005608752253465354
step: 170, loss: 0.000308086076984182
step: 180, loss: 0.00033281833748333156
step: 190, loss: 0.00038449998828582466
step: 200, loss: 0.0009025047183968127
step: 210, loss: 0.00010656254744390026
step: 220, loss: 0.0002608874347060919
step: 230, loss: 0.029710419476032257
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9876819708846584, f1=0.9764309764309763,
best_f1=0.9853438556933484
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02921650931239128
step: 10, loss: 0.03411230072379112
step: 20, loss: 0.010329940356314182
step: 30, loss: 0.00023209872597362846
step: 40, loss: 7.634820940438658e-05
step: 50, loss: 0.00011443792027421296
step: 60, loss: 0.00015781780530232936
step: 70, loss: 0.00021473605011124164
step: 80, loss: 0.0012699951184913516
step: 90, loss: 0.00012957035505678505
step: 100, loss: 0.00018814609211403877
step: 110, loss: 0.00024920201394706964
step: 120, loss: 0.00020104854775127023
step: 130, loss: 0.008863233029842377
step: 140, loss: 0.12083804607391357
step: 150, loss: 9.207244875142351e-05
step: 160, loss: 0.00015114180860109627
step: 170, loss: 0.0004042026412207633
step: 180, loss: 0.005081820767372847
step: 190, loss: 0.0011262381449341774
step: 200, loss: 0.002021027496084571
step: 210, loss: 0.00015491279191337526
step: 220, loss: 0.0003035369445569813
step: 230, loss: 0.039614222943782806
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9807037457434733, f1=0.9761634506242906,
best_f1=0.9853438556933484
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.00023466798302251846
step: 10, loss: 0.0001299870345974341
step: 20, loss: 0.03010079264640808
step: 30, loss: 0.0030916945543140173
step: 40, loss: 0.00010110576840816066
step: 50, loss: 0.00012572291598189622
step: 60, loss: 0.00047371070832014084
step: 70, loss: 0.00035759113961830735
step: 80, loss: 0.04583638533949852
step: 90, loss: 0.0010402232874184847
step: 100, loss: 0.002568027935922146
step: 110, loss: 0.00010116649355040863
step: 120, loss: 0.03507062792778015
step: 130, loss: 0.00025469952379353344
step: 140, loss: 0.014798465184867382
step: 150, loss: 0.0021552988328039646
step: 160, loss: 0.0011094107758253813
step: 170, loss: 0.00026225915644317865
step: 180, loss: 0.00043004515464417636
step: 190, loss: 8.084123692242429e-05
step: 200, loss: 0.00040450418600812554
step: 210, loss: 0.00017789701814763248
step: 220, loss: 0.03725713863968849
step: 230, loss: 0.0058769891038537025
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9864559819413092, f1=0.9784335981838819,
best_f1=0.9853438556933484
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001682690199231729
step: 10, loss: 0.0002688764070626348
step: 20, loss: 0.0003791056224144995
step: 30, loss: 0.00012968305964022875
step: 40, loss: 5.348698687157594e-05
step: 50, loss: 0.005657610949128866
step: 60, loss: 0.028456125408411026
step: 70, loss: 0.0001772919058566913
step: 80, loss: 0.00038352832780219615
step: 90, loss: 0.00019548498676158488
step: 100, loss: 0.00045029554166831076
step: 110, loss: 0.00018731212185230106
step: 120, loss: 0.021232573315501213
step: 130, loss: 0.0005093974177725613
step: 140, loss: 0.026444952934980392
step: 150, loss: 0.00010396289872005582

```

```

step: 160, loss: 0.00012689331197179854
step: 170, loss: 0.0015212107682600617
step: 180, loss: 0.0002484995638951659
step: 190, loss: 0.00013830498210154474
step: 200, loss: 6.904520705575123e-05
step: 210, loss: 0.0005840190569870174
step: 220, loss: 0.00017345597734674811
step: 230, loss: 0.000827991811092943
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9842342342342343, f1=0.9775280898876404,
best_f1=0.9853438556933484
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.516788850305602e-05
step: 10, loss: 0.00012796660303138196
step: 20, loss: 3.871980879921466e-05
step: 30, loss: 0.00011829142749775201
step: 40, loss: 0.011652503162622452
step: 50, loss: 0.001575962989591062
step: 60, loss: 9.336850780528039e-05
step: 70, loss: 0.00018701827502809465
step: 80, loss: 0.00019768084166571498
step: 90, loss: 5.689742465619929e-05
step: 100, loss: 0.0003874096437357366
step: 110, loss: 9.159867477137595e-05
step: 120, loss: 7.20925017958507e-05
step: 130, loss: 0.00036391004687175155
step: 140, loss: 6.233858584892005e-05
step: 150, loss: 9.242330270353705e-05
step: 160, loss: 6.949118687771261e-05
step: 170, loss: 0.010522185824811459
step: 180, loss: 0.0231672041118145
step: 190, loss: 9.56410585786216e-05
step: 200, loss: 0.00011519736290210858
step: 210, loss: 5.709530887543224e-05
step: 220, loss: 7.07618091837503e-05
step: 230, loss: 0.019480014219880104
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9887133182844244, f1=0.9807474518686297,
best_f1=0.9853438556933484
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 9.250283619621769e-05
step: 10, loss: 0.02408815175294876
step: 20, loss: 6.141015910543501e-05
step: 30, loss: 6.663402018602937e-05
step: 40, loss: 6.15466560702771e-05
step: 50, loss: 0.00010024916991824284
step: 60, loss: 0.015427358448505402
step: 70, loss: 0.00012013183732051402
step: 80, loss: 4.8881411203183234e-05
step: 90, loss: 4.9489110097056255e-05
step: 100, loss: 5.1972296205349267e-05
step: 110, loss: 0.0003604409284889698
step: 120, loss: 9.413881343789399e-05
step: 130, loss: 7.912788714747876e-05
step: 140, loss: 0.00010263115109410137
step: 150, loss: 2.3308333766181022e-05
step: 160, loss: 0.0407039113342762
step: 170, loss: 7.713428203715011e-05
step: 180, loss: 4.3876931158592924e-05
step: 190, loss: 7.836770237190649e-05
step: 200, loss: 0.007264596410095692
step: 210, loss: 9.058423165697604e-05
step: 220, loss: 0.024659516289830208
step: 230, loss: 6.357781239785254e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9875706214689265, f1=0.979591836734694,
best_f1=0.9853438556933484
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00031487265368923545
step: 10, loss: 0.00019912509014829993
step: 20, loss: 0.00010927164112217724
step: 30, loss: 0.026310987770557404
step: 40, loss: 0.0002876987273339182
step: 50, loss: 0.0021697019692510366
step: 60, loss: 8.993173105409369e-05
step: 70, loss: 7.854164869058877e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 80, loss: 0.0004312090459279716
step: 90, loss: 0.0006903823232278228
step: 100, loss: 0.00019391679961699992
step: 110, loss: 0.00014492595801129937
step: 120, loss: 6.464743637479842e-05
step: 130, loss: 8.010634337551892e-05
step: 140, loss: 0.04395042359828949

```



```

step: 150, loss: 0.024496518075466156
step: 160, loss: 4.5068027247907594e-05
step: 170, loss: 0.00010677078535081819
step: 180, loss: 0.0005795189063064754
step: 190, loss: 3.402510628802702e-05
step: 200, loss: 7.394195563392714e-05
step: 210, loss: 0.00010012404527515173
step: 220, loss: 2.6914080081041902e-05
step: 230, loss: 5.0284696044400334e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9887133182844244, f1=0.9819413092550789,
best_f1=0.9853438556933484
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00010514556197449565
step: 10, loss: 8.233392145484686e-05
step: 20, loss: 0.013299195095896721
step: 30, loss: 6.459834548877552e-05
step: 40, loss: 0.00022888365492690355
step: 50, loss: 6.045719419489615e-05
step: 60, loss: 3.562855636118911e-05
step: 70, loss: 5.4078227549325675e-05
step: 80, loss: 9.169690747512504e-05
step: 90, loss: 0.0008538747788406909
step: 100, loss: 0.01849646307528019
step: 110, loss: 6.903470057295635e-05
step: 120, loss: 7.815269782440737e-05
step: 130, loss: 6.75266855978407e-05
step: 140, loss: 5.2576044254237786e-05
step: 150, loss: 0.0001005012818495743
step: 160, loss: 2.8400168957887217e-05
step: 170, loss: 3.9585578633705154e-05
step: 180, loss: 8.758678450249135e-05
step: 190, loss: 0.00045508486800827086
step: 200, loss: 3.588746039895341e-05
step: 210, loss: 0.00022654622443951666
step: 220, loss: 0.0001413743302691728
step: 230, loss: 1.973610596905928e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9887640449438202, f1=0.980963045912654,
best_f1=0.9853438556933484
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005570637993514538
step: 10, loss: 5.704384602722712e-05
step: 20, loss: 6.18768681306392e-05
step: 30, loss: 0.0001248182961717248
step: 40, loss: 6.593047874048352e-05
step: 50, loss: 7.146179268602282e-05
step: 60, loss: 7.777456630719826e-05
step: 70, loss: 7.5208256021142e-05
step: 80, loss: 0.019025912508368492
step: 90, loss: 3.004626341862604e-05
step: 100, loss: 6.35658871033229e-05
step: 110, loss: 5.32150334038306e-05
step: 120, loss: 0.00020969459728803486
step: 130, loss: 6.730936001986265e-05
step: 140, loss: 8.495830843457952e-05
step: 150, loss: 0.0001551073946757242
step: 160, loss: 5.09604396938812e-05
step: 170, loss: 3.965937503380701e-05
step: 180, loss: 7.729642675258219e-05
step: 190, loss: 0.018821939826011658
step: 200, loss: 5.576650073635392e-05
step: 210, loss: 0.027975203469395638
step: 220, loss: 0.00013184279669076204
step: 230, loss: 8.915408398024738e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9887640449438202, f1=0.9809203142536477,
best_f1=0.9853438556933484
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False

```

```

keep_batchnorm_fp32      : True
master_weights           : True
loss_scale               : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2473it [00:08, 282.54it/s]
load_f1 = 0.9854096520763187
real_f1 = 0.9865470852017937
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:12, 364.28it/s]

```

1.1.3 DBLP-GoogleScholar - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
    --task Structured/DBLP-GoogleScholar \
    --batch_size 32 \
    --max_len 256 \
    --lr 3e-5 \
    --n_epochs 15 \
    --finetuning \
    --lm distilbert \
    --fp16 \
    --save_model \
    --da entry_swap \
    --da span_shuffle \
    --summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-GoogleScholar.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
    --task Structured/DBLP-GoogleScholar \

```

```

--input_path input/input_small.jsonl \
--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/

```

Downloading: 100% 442/442 [00:00<00:00, 387kB/s]
 Downloading: 100% 232k/232k [00:00<00:00, 3.25MB/s]
 Downloading: 100% 268M/268M [00:04<00:00, 66.3MB/s]
 Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.

Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

warnings.warn("An input tensor was not cuda.")

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:

UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order:

`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate", UserWarning)

step: 0, loss: 0.7961419820785522

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0

step: 10, loss: 0.45647603273391724
step: 20, loss: 0.4908207356929779
step: 30, loss: 0.42812490463256836
step: 40, loss: 0.3274894058704376
step: 50, loss: 0.1796170324087143
step: 60, loss: 0.1418589949607849
step: 70, loss: 0.13013866543769836
step: 80, loss: 0.18564088642597198
step: 90, loss: 0.21104983985424042
step: 100, loss: 0.2992580831050873
step: 110, loss: 0.08439783751964569
step: 120, loss: 0.060572993010282516
step: 130, loss: 0.012844391167163849
step: 140, loss: 0.335885226726532
step: 150, loss: 0.05772220715880394
step: 160, loss: 0.13802441954612732
step: 170, loss: 0.1422475427389145
step: 180, loss: 0.1370546668767929
step: 190, loss: 0.10652749985456467
step: 200, loss: 0.07607006281614304
step: 210, loss: 0.06658700108528137
step: 220, loss: 0.03776880353689194
step: 230, loss: 0.17846791446208954
step: 240, loss: 0.10776084661483765
step: 250, loss: 0.07252278923988342
step: 260, loss: 0.010961239226162434
step: 270, loss: 0.011269800364971161
step: 280, loss: 0.10030075162649155
step: 290, loss: 0.07170161604881287
step: 300, loss: 0.1519356667995453
step: 310, loss: 0.06187528744339943
step: 320, loss: 0.14053110778331757
step: 330, loss: 0.07342492043972015
step: 340, loss: 0.2007521092891693
step: 350, loss: 0.12340915948152542
step: 360, loss: 0.035048067569732666
step: 370, loss: 0.14438088238239288
step: 380, loss: 0.1377612054347992
step: 390, loss: 0.03786597028374672
step: 400, loss: 0.013393363915383816
step: 410, loss: 0.06107799708843231
step: 420, loss: 0.007706951815634966
step: 430, loss: 0.11280304938554764
step: 440, loss: 0.13540609180927277
step: 450, loss: 0.02178211882710457
step: 460, loss: 0.04445364326238632
step: 470, loss: 0.28272920846939087

```

step: 480, loss: 0.2491627335548401
step: 490, loss: 0.03914802521467209
step: 500, loss: 0.005119688808917999
step: 510, loss: 0.034291718155145645
step: 520, loss: 0.11352105438709259
step: 530, loss: 0.18205171823501587
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9311740890688259, f1=0.9180474697716078,
best_f1=0.9180474697716078
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.19534114003181458
step: 10, loss: 0.14375466108322144
step: 20, loss: 0.04594849795103073
step: 30, loss: 0.04690760374069214
step: 40, loss: 0.0037918256130069494
step: 50, loss: 0.1670338213443756
step: 60, loss: 0.2632628381252289
step: 70, loss: 0.03856688365340233
step: 80, loss: 0.008300332352519035
step: 90, loss: 0.005436813924461603
step: 100, loss: 0.4228407144546509
step: 110, loss: 0.03887595236301422
step: 120, loss: 0.09509137272834778
step: 130, loss: 0.061828356236219406
step: 140, loss: 0.03594006225466728
step: 150, loss: 0.05860590189695358
step: 160, loss: 0.008291070349514484
step: 170, loss: 0.02217838168144226
step: 180, loss: 0.05647333338856697
step: 190, loss: 0.02727455645799637
step: 200, loss: 0.01923990435898304
step: 210, loss: 0.03107992559671402
step: 220, loss: 0.024012157693505287
step: 230, loss: 0.01354643888771534
step: 240, loss: 0.1219547688961029
step: 250, loss: 0.03630907088518143
step: 260, loss: 0.020894691348075867
step: 270, loss: 0.10778965801000595
step: 280, loss: 0.2713981568813324
step: 290, loss: 0.09718194603919983

```

```

step: 300, loss: 0.031669408082962036
step: 310, loss: 0.08500134944915771
step: 320, loss: 0.09755872935056686
step: 330, loss: 0.07663017511367798
step: 340, loss: 0.004966039676219225
step: 350, loss: 0.050561919808387756
step: 360, loss: 0.05665656179189682
step: 370, loss: 0.02838900499045849
step: 380, loss: 0.0568368174135685
step: 390, loss: 0.04221951216459274
step: 400, loss: 0.0504876971244812
step: 410, loss: 0.0006419049459509552
step: 420, loss: 0.018922675400972366
step: 430, loss: 0.01290992833673954
step: 440, loss: 0.028019804507493973
step: 450, loss: 0.014612923376262188
step: 460, loss: 0.2229411005973816
step: 470, loss: 0.045070476830005646
step: 480, loss: 0.2889871299266815
step: 490, loss: 0.0500003956258297
step: 500, loss: 0.006364055909216404
step: 510, loss: 0.04379640147089958
step: 520, loss: 0.06740912050008774
step: 530, loss: 0.18793995678424835
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9396035039188566, f1=0.9275229357798166,
best_f1=0.9275229357798166
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07332415133714676
step: 10, loss: 0.04729143902659416
step: 20, loss: 0.1351131647825241
step: 30, loss: 0.21014820039272308
step: 40, loss: 0.002624575747177005
step: 50, loss: 0.006210759282112122
step: 60, loss: 0.003958650398999453
step: 70, loss: 0.06639177352190018
step: 80, loss: 0.0007429186953231692
step: 90, loss: 0.1464449167251587
step: 100, loss: 0.038499198853969574
step: 110, loss: 0.07543718814849854

```

```
step: 120, loss: 0.009979212656617165
step: 130, loss: 0.04676763340830803
step: 140, loss: 0.01921018399298191
step: 150, loss: 0.017922453582286835
step: 160, loss: 0.013135280460119247
step: 170, loss: 0.002862471854314208
step: 180, loss: 0.00439068116247654
step: 190, loss: 0.0045450902543962
step: 200, loss: 0.005972580052912235
step: 210, loss: 0.05129546672105789
step: 220, loss: 0.03779711574316025
step: 230, loss: 0.028406767174601555
step: 240, loss: 0.006244391202926636
step: 250, loss: 0.00709448242560029
step: 260, loss: 0.014105540700256824
step: 270, loss: 0.0013271818170323968
step: 280, loss: 0.014251673594117165
step: 290, loss: 0.023966079577803612
step: 300, loss: 0.11228283494710922
step: 310, loss: 0.17930930852890015
step: 320, loss: 0.1675615906715393
step: 330, loss: 0.00781447347253561
step: 340, loss: 0.0020076953805983067
step: 350, loss: 0.039206407964229584
step: 360, loss: 0.02849496901035309
step: 370, loss: 0.004266458563506603
step: 380, loss: 0.023536723107099533
step: 390, loss: 0.026209695264697075
step: 400, loss: 0.0021050870418548584
step: 410, loss: 0.0042877197265625
step: 420, loss: 0.059543490409851074
step: 430, loss: 0.10104464739561081
step: 440, loss: 0.011929205618798733
step: 450, loss: 0.048238132148981094
step: 460, loss: 0.023405203595757484
step: 470, loss: 0.0015993815613910556
step: 480, loss: 0.0020386590622365475
step: 490, loss: 0.10916459560394287
step: 500, loss: 0.12698984146118164
step: 510, loss: 0.019073152914643288
step: 520, loss: 0.008645494468510151
step: 530, loss: 0.014269850216805935
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.9338959212376934, f1=0.9245020842982862,
best_f1=0.9275229357798166
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```


An input tensor was not cuda.

warnings.warn("An input tensor was not cuda.")

step: 0, loss: 0.005960258189588785
step: 10, loss: 0.009282749146223068
step: 20, loss: 0.008769736625254154
step: 30, loss: 0.048210710287094116
step: 40, loss: 0.001528145861811936
step: 50, loss: 0.025759847834706306
step: 60, loss: 0.0013117427006363869
step: 70, loss: 0.003038091119378805
step: 80, loss: 0.20257066190242767
step: 90, loss: 0.001546404673717916
step: 100, loss: 0.08255505561828613
step: 110, loss: 0.0028381687588989735
step: 120, loss: 0.006004479248076677
step: 130, loss: 0.029226602986454964
step: 140, loss: 0.012132660485804081
step: 150, loss: 0.0018439647974446416
step: 160, loss: 0.014892224222421646
step: 170, loss: 0.0038013560697436333
step: 180, loss: 0.1257801353931427
step: 190, loss: 0.03001600131392479
step: 200, loss: 0.0004168977029621601
step: 210, loss: 0.04203035682439804
step: 220, loss: 0.003050441388040781
step: 230, loss: 0.16413487493991852
step: 240, loss: 0.026699386537075043
step: 250, loss: 0.003013728419318795
step: 260, loss: 0.14796055853366852
step: 270, loss: 0.005414216313511133
step: 280, loss: 0.001080243499018252
step: 290, loss: 0.09992441534996033
step: 300, loss: 0.0036472557112574577
step: 310, loss: 0.03032965213060379
step: 320, loss: 0.06506338715553284
step: 330, loss: 0.01310780644416809
step: 340, loss: 0.04102065786719322
step: 350, loss: 0.001715941121801734
step: 360, loss: 0.0013525496469810605
step: 370, loss: 0.12793011963367462
step: 380, loss: 0.0047850473783910275
step: 390, loss: 0.023769477382302284
step: 400, loss: 0.035957060754299164
step: 410, loss: 0.01136774942278862
step: 420, loss: 0.0037565322127193213
step: 430, loss: 0.006313383113592863
step: 440, loss: 0.03584180772304535
step: 450, loss: 0.004873095545917749

```

step: 460, loss: 0.0012197081232443452
step: 470, loss: 0.03348579257726669
step: 480, loss: 0.006821559276431799
step: 490, loss: 0.12709742784500122
step: 500, loss: 0.008765807375311852
step: 510, loss: 0.016284596174955368
step: 520, loss: 0.00243604788556695
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 530, loss: 0.002618612488731742
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.9397930385700847, f1=0.9265799256505577,
best_f1=0.9265799256505577
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0015782337868586183
step: 10, loss: 0.021559886634349823
step: 20, loss: 0.0077919745817780495
step: 30, loss: 0.0038181182462722063
step: 40, loss: 0.024585476145148277
step: 50, loss: 0.0018426107708364725
step: 60, loss: 0.0008765594684518874
step: 70, loss: 0.0002460329851601273
step: 80, loss: 0.008414966985583305
step: 90, loss: 0.056702349334955215
step: 100, loss: 0.0004020882479380816
step: 110, loss: 0.0014078090898692608
step: 120, loss: 0.0009613430593162775
step: 130, loss: 0.000524252827744931
step: 140, loss: 0.0020799485500901937
step: 150, loss: 0.014780052937567234
step: 160, loss: 0.21900029480457306
step: 170, loss: 0.0011343235382810235
step: 180, loss: 0.002030411036685109
step: 190, loss: 0.002587713534012437
step: 200, loss: 0.01761792227625847
step: 210, loss: 0.0012090080417692661
step: 220, loss: 0.014408564195036888
step: 230, loss: 0.0029231125954538584
step: 240, loss: 0.003772921394556761
step: 250, loss: 0.0001622449344722554
step: 260, loss: 0.0007833386771380901

```

```

step: 270, loss: 0.0030145354103296995
step: 280, loss: 0.07610326260328293
step: 290, loss: 0.001255506300367415
step: 300, loss: 0.003273677546530962
step: 310, loss: 0.0019908472895622253
step: 320, loss: 0.009097612462937832
step: 330, loss: 0.0037545461673289537
step: 340, loss: 0.016831083223223686
step: 350, loss: 0.006250783801078796
step: 360, loss: 0.0009456214029341936
step: 370, loss: 0.0010684990556910634
step: 380, loss: 0.020688051357865334
step: 390, loss: 0.011681183241307735
step: 400, loss: 0.021410491317510605
step: 410, loss: 0.0038583585992455482
step: 420, loss: 0.002643611514940858
step: 430, loss: 0.054294586181640625
step: 440, loss: 0.013391218148171902
step: 450, loss: 0.024796053767204285
step: 460, loss: 0.005927113816142082
step: 470, loss: 0.07498525828123093
step: 480, loss: 0.0012561656767502427
step: 490, loss: 0.054393380880355835
step: 500, loss: 0.0029002889059484005
step: 510, loss: 0.003942691721022129
step: 520, loss: 0.0010292925871908665
step: 530, loss: 0.012825911864638329
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.938134810710988, f1=0.9316317228805834,
best_f1=0.9265799256505577
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008368728449568152
step: 10, loss: 0.004943004809319973
step: 20, loss: 0.00046570698032155633
step: 30, loss: 0.008807732723653316
step: 40, loss: 0.021051742136478424
step: 50, loss: 0.0009218530030921102
step: 60, loss: 0.0007429799297824502
step: 70, loss: 0.013304118998348713
step: 80, loss: 0.00043262584949843585
step: 90, loss: 0.01979953609406948
step: 100, loss: 0.07174176722764969
step: 110, loss: 0.002312222495675087
step: 120, loss: 0.0005445326678454876

```

```

step: 130, loss: 0.0009350342443212867
step: 140, loss: 0.0008466222789138556
step: 150, loss: 0.00575594836845994
step: 160, loss: 0.001119619351811707
step: 170, loss: 0.0002200048475060612
step: 180, loss: 0.001692634541541338
step: 190, loss: 0.0011932568158954382
step: 200, loss: 0.00025044105132110417
step: 210, loss: 0.0009756796061992645
step: 220, loss: 0.0004273473168723285
step: 230, loss: 0.0018628101097419858
step: 240, loss: 0.006414234172552824
step: 250, loss: 0.00025262878625653684
step: 260, loss: 0.002683299360796809
step: 270, loss: 0.0012174822622910142
step: 280, loss: 0.01682831160724163
step: 290, loss: 0.004681578371673822
step: 300, loss: 0.0002908124588429928
step: 310, loss: 0.00033179234014824033
step: 320, loss: 0.020958542823791504
step: 330, loss: 0.0006760098622180521
step: 340, loss: 0.000737756141461432
step: 350, loss: 0.09067428112030029
step: 360, loss: 0.01695326343178749
step: 370, loss: 0.001970818731933832
step: 380, loss: 0.0582512803375721
step: 390, loss: 0.01245139166712761
step: 400, loss: 7.720165012869984e-05
step: 410, loss: 0.00011836893827421591
step: 420, loss: 0.036168284714221954
step: 430, loss: 0.06596130132675171
step: 440, loss: 0.0010804567718878388
step: 450, loss: 0.0007743701571598649
step: 460, loss: 0.0020956741645932198
step: 470, loss: 0.12296763062477112
step: 480, loss: 0.018539557233452797
step: 490, loss: 0.0002906701120082289
step: 500, loss: 0.0016656641382724047
step: 510, loss: 0.00022584351245313883
step: 520, loss: 0.0016017109155654907
step: 530, loss: 0.0006417374825105071
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.9283720930232557, f1=0.9245107176141659,
best_f1=0.9265799256505577
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.017998160794377327
step: 10, loss: 0.016004478558897972
step: 20, loss: 0.002753800479695201
step: 30, loss: 0.004647867754101753
step: 40, loss: 0.0006404724554158747
step: 50, loss: 0.07176414132118225
step: 60, loss: 0.0002887934970203787
step: 70, loss: 0.05923491716384888
step: 80, loss: 0.00517965666949749
step: 90, loss: 0.007943261414766312
step: 100, loss: 0.0017443244578316808
step: 110, loss: 0.0007574431947432458
step: 120, loss: 9.685867553343996e-05
step: 130, loss: 0.0012019699206575751
step: 140, loss: 0.0005274272407405078
step: 150, loss: 8.859393710736185e-05
step: 160, loss: 0.0005408000433817506
step: 170, loss: 0.0007794983685016632
step: 180, loss: 0.0059726787731051445
step: 190, loss: 0.00014154688687995076
step: 200, loss: 0.0019179022638127208
step: 210, loss: 0.0006610325071960688
step: 220, loss: 6.789275357732549e-05
step: 230, loss: 0.0006495968555100262
step: 240, loss: 0.0024159951135516167
step: 250, loss: 0.0002831253514159471
step: 260, loss: 0.0035139445681124926
step: 270, loss: 0.0001425655500497669
step: 280, loss: 0.09064195305109024
step: 290, loss: 0.0019294261001050472
step: 300, loss: 0.0008190071675926447
step: 310, loss: 0.0013741422444581985
step: 320, loss: 0.026980087161064148
step: 330, loss: 0.0003768657916225493
step: 340, loss: 0.0029882132075726986
step: 350, loss: 0.0004023822257295251
step: 360, loss: 0.0012303460389375687
step: 370, loss: 0.0013999136863276362
step: 380, loss: 0.0031722683925181627
step: 390, loss: 0.00016029032121878117
step: 400, loss: 4.114425973966718e-05
step: 410, loss: 0.005670797545462847
step: 420, loss: 0.00013217238301876932
step: 430, loss: 0.00019090886053163558
step: 440, loss: 0.00039822241524234414
step: 450, loss: 0.0042926534079015255
step: 460, loss: 0.001804499072022736
```

```

step: 470, loss: 0.005997100844979286
step: 480, loss: 0.0024501935113221407
step: 490, loss: 0.0005215308046899736
step: 500, loss: 8.692628762219101e-05
step: 510, loss: 0.0007577587966807187
step: 520, loss: 0.00010696699609979987
step: 530, loss: 0.0002614464610815048
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9340148698884759, f1=0.9286376274328081,
best_f1=0.9265799256505577
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009943179320544004
step: 10, loss: 0.003501179162412882
step: 20, loss: 0.03483262285590172
step: 30, loss: 0.0019715602975338697
step: 40, loss: 0.0009155316511169076
step: 50, loss: 0.0012215453898534179
step: 60, loss: 0.0007694161031395197
step: 70, loss: 0.0015177028253674507
step: 80, loss: 0.00014866686251480132
step: 90, loss: 0.002722493838518858
step: 100, loss: 0.04285908117890358
step: 110, loss: 0.00021682537044398487
step: 120, loss: 0.0018103616312146187
step: 130, loss: 0.0015767368022352457
step: 140, loss: 3.9530048525193706e-05
step: 150, loss: 0.022568490356206894
step: 160, loss: 0.001453959266655147
step: 170, loss: 0.0021459138952195644
step: 180, loss: 0.005607393570244312
step: 190, loss: 0.009288031607866287
step: 200, loss: 0.00039107241900637746
step: 210, loss: 0.00028469046810641885
step: 220, loss: 0.0010960606159642339
step: 230, loss: 0.00019855254504363984
step: 240, loss: 0.0001345290511380881
step: 250, loss: 0.0001159329476649873
step: 260, loss: 0.0017342306673526764
step: 270, loss: 4.24221798311919e-05
step: 280, loss: 0.0002906814916059375
step: 290, loss: 0.0020493236370384693
step: 300, loss: 8.31064535304904e-05
step: 310, loss: 0.01745201274752617
step: 320, loss: 0.002811640501022339

```

```

step: 330, loss: 0.0062307873740792274
step: 340, loss: 0.003994342405349016
step: 350, loss: 0.00159797677770257
step: 360, loss: 0.0007123834220692515
step: 370, loss: 0.00022583561076316983
step: 380, loss: 0.00015298393554985523
step: 390, loss: 0.09758234769105911
step: 400, loss: 0.11727333068847656
step: 410, loss: 0.00039232848212122917
step: 420, loss: 0.00015804925351403654
step: 430, loss: 0.00035085377749055624
step: 440, loss: 0.02790920063853264
step: 450, loss: 0.0011768051190301776
step: 460, loss: 0.00019812579557765275
step: 470, loss: 0.00011439322406658903
step: 480, loss: 7.214963261503726e-05
step: 490, loss: 0.008598500862717628
step: 500, loss: 3.690449375426397e-05
step: 510, loss: 2.8803136956412345e-05
step: 520, loss: 7.325187470996752e-05
step: 530, loss: 0.00013500847853720188
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9343955014058105, f1=0.9312762973352035,
best_f1=0.9265799256505577
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.22213148442097e-05
step: 10, loss: 3.206238761777058e-05
step: 20, loss: 0.00038218245026655495
step: 30, loss: 0.0011853374307975173
step: 40, loss: 0.0003410396457184106
step: 50, loss: 0.0005309861735440791
step: 60, loss: 0.0002747373073361814
step: 70, loss: 0.11933476477861404
step: 80, loss: 0.00010501259384909645
step: 90, loss: 0.00024599311291240156
step: 100, loss: 0.0003088050289079547
step: 110, loss: 0.0005496959784068167
step: 120, loss: 0.004336580168455839
step: 130, loss: 0.00037957701715640724
step: 140, loss: 0.0005898473900742829
step: 150, loss: 0.00016229518223553896
step: 160, loss: 0.002118384465575218
step: 170, loss: 0.0005743824294768274
step: 180, loss: 0.00062564667314291

```

```

step: 190, loss: 0.0013659547548741102
step: 200, loss: 0.0008423744002357125
step: 210, loss: 3.83492952096276e-05
step: 220, loss: 0.000980912707746029
step: 230, loss: 4.4855911255581304e-05
step: 240, loss: 0.00011034664930775762
step: 250, loss: 0.008420220576226711
step: 260, loss: 0.0003022592281922698
step: 270, loss: 0.00015692831948399544
step: 280, loss: 3.3135231205960736e-05
step: 290, loss: 0.00012398180842865258
step: 300, loss: 0.0003586452512536198
step: 310, loss: 2.694421709747985e-05
step: 320, loss: 0.0006463729660026729
step: 330, loss: 4.4891236029798165e-05
step: 340, loss: 5.456954750115983e-05
step: 350, loss: 3.533716517267749e-05
step: 360, loss: 0.018615614622831345
step: 370, loss: 0.00011586388427531347
step: 380, loss: 0.00028173092869110405
step: 390, loss: 0.00016625755233690143
step: 400, loss: 0.007038231939077377
step: 410, loss: 6.789878534618765e-05
step: 420, loss: 8.071713818935677e-05
step: 430, loss: 2.9458096832968295e-05
step: 440, loss: 0.00048624275950714946
step: 450, loss: 0.00018203252693638206
step: 460, loss: 0.00035788241075351834
step: 470, loss: 9.082420729100704e-05
step: 480, loss: 0.0004708743072114885
step: 490, loss: 0.0009173275320790708
step: 500, loss: 0.0014356172177940607
step: 510, loss: 0.004054485820233822
step: 520, loss: 0.0009727359283715487
step: 530, loss: 0.00017036049393936992
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9331450094161958, f1=0.9315960912052118,
best_f1=0.9265799256505577
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004336446290835738
step: 10, loss: 0.00012561881158035249
step: 20, loss: 6.180407217470929e-05
step: 30, loss: 6.220828072400764e-05
step: 40, loss: 7.063943485263735e-05

```


step: 50, loss: 0.00014810659922659397
step: 60, loss: 6.65795014356263e-05
step: 70, loss: 6.767256127204746e-05
step: 80, loss: 5.313153815222904e-05
step: 90, loss: 0.00013438369205687195
step: 100, loss: 0.003221443621441722
step: 110, loss: 0.0001842817582655698
step: 120, loss: 3.6747194826602936e-05
step: 130, loss: 0.00015982547483872622
step: 140, loss: 6.488712824648246e-05
step: 150, loss: 3.2563857530476525e-05
step: 160, loss: 0.00029392566648311913
step: 170, loss: 0.0001609566097613424
step: 180, loss: 0.00016134623729158193
step: 190, loss: 9.742324618855491e-05
step: 200, loss: 0.00033550054649822414
step: 210, loss: 3.2937514333752915e-05
step: 220, loss: 2.5070759875234216e-05
step: 230, loss: 3.4412802051519975e-05
step: 240, loss: 2.5245821234420873e-05
step: 250, loss: 0.00010110639414051548
step: 260, loss: 0.00041350393439643085
step: 270, loss: 0.0005788019625470042
step: 280, loss: 0.005721925292164087
step: 290, loss: 2.8955937523278408e-05
step: 300, loss: 7.526706758653745e-05
step: 310, loss: 8.698049350641668e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 320, loss: 0.00016887632955331355
step: 330, loss: 0.013406902551651001
step: 340, loss: 5.668766607414e-05
step: 350, loss: 2.901157677115407e-05
step: 360, loss: 2.4828550522215664e-05
step: 370, loss: 7.885607919888571e-05
step: 380, loss: 4.919238563161343e-05
step: 390, loss: 4.642323619918898e-05
step: 400, loss: 2.549152122810483e-05
step: 410, loss: 9.97089664451778e-05
step: 420, loss: 0.0005397195927798748
step: 430, loss: 2.8329986889730208e-05
step: 440, loss: 2.709337422857061e-05
step: 450, loss: 0.0036771653685718775
step: 460, loss: 4.365672793937847e-05
step: 470, loss: 2.9253747925395146e-05
step: 480, loss: 4.46711819677148e-05
step: 490, loss: 0.018554218113422394
step: 500, loss: 0.0013246331363916397
step: 510, loss: 0.0002196420100517571

```

step: 520, loss: 0.005595542956143618
step: 530, loss: 0.002081114100292325
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9275084013442151, f1=0.9198271723475756,
best_f1=0.9265799256505577
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00021737211500294507
step: 10, loss: 0.001846211845986545
step: 20, loss: 0.0003214836760889739
step: 30, loss: 0.00048602386959828436
step: 40, loss: 0.19207105040550232
step: 50, loss: 0.0002525532909203321
step: 60, loss: 5.278474054648541e-05
step: 70, loss: 0.0012980173341929913
step: 80, loss: 0.000685092352796346
step: 90, loss: 0.00018187673413194716
step: 100, loss: 3.632691004895605e-05
step: 110, loss: 9.50276298681274e-05
step: 120, loss: 0.0004008929245173931
step: 130, loss: 0.00034002031316049397
step: 140, loss: 3.453872704994865e-05
step: 150, loss: 3.7738383980467916e-05
step: 160, loss: 0.0037622342351824045
step: 170, loss: 7.748697680654004e-05
step: 180, loss: 3.5417589970165864e-05
step: 190, loss: 0.0004883302026428282
step: 200, loss: 8.656905993120745e-05
step: 210, loss: 8.613879617769271e-05
step: 220, loss: 4.897664621239528e-05
step: 230, loss: 5.727592724724673e-05
step: 240, loss: 3.0311095542856492e-05
step: 250, loss: 3.189098060829565e-05
step: 260, loss: 2.9052533136564307e-05
step: 270, loss: 0.0002921954437624663
step: 280, loss: 5.731168857892044e-05
step: 290, loss: 0.022584816440939903
step: 300, loss: 0.0005533098010346293
step: 310, loss: 0.00035913533065468073
step: 320, loss: 0.015454034321010113
step: 330, loss: 3.590837513911538e-05
step: 340, loss: 0.0018801761325448751
step: 350, loss: 0.011534790508449078
step: 360, loss: 0.00031888685771264136
step: 370, loss: 8.229199011111632e-05

```

```

step: 380, loss: 3.379174813744612e-05
step: 390, loss: 0.02572794258594513
step: 400, loss: 2.5372193704242818e-05
step: 410, loss: 0.0024057175032794476
step: 420, loss: 0.01839522458612919
step: 430, loss: 0.007795283570885658
step: 440, loss: 7.836265285732225e-05
step: 450, loss: 2.18744644371327e-05
step: 460, loss: 0.0023583511356264353
step: 470, loss: 0.00035566126462072134
step: 480, loss: 4.0274986531585455e-05
step: 490, loss: 0.001534582581371069
step: 500, loss: 0.00019607525609899312
step: 510, loss: 4.2677002056734636e-05
step: 520, loss: 2.7037413019570522e-05
step: 530, loss: 6.938911974430084e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9374416433239963, f1=0.9288040949278734,
best_f1=0.9265799256505577
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000949012755881995
step: 10, loss: 8.652736141812056e-05
step: 20, loss: 3.07349837385118e-05
step: 30, loss: 4.9799007683759555e-05
step: 40, loss: 8.652736141812056e-05
step: 50, loss: 0.00016789609799161553
step: 60, loss: 3.391726204426959e-05
step: 70, loss: 0.00015819576219655573
step: 80, loss: 3.0676586902700365e-05
step: 90, loss: 3.160386040690355e-05
step: 100, loss: 4.034025187138468e-05
step: 110, loss: 2.6780164262163453e-05
step: 120, loss: 0.03720148652791977
step: 130, loss: 3.074708001804538e-05
step: 140, loss: 0.00011275552242295817
step: 150, loss: 0.0001413449936080724
step: 160, loss: 5.8167170209344476e-05
step: 170, loss: 2.3088805392035283e-05
step: 180, loss: 3.7630092265317217e-05
step: 190, loss: 2.4839358957251534e-05
step: 200, loss: 2.8109983759350143e-05
step: 210, loss: 0.0008563963347114623
step: 220, loss: 3.78969925804995e-05
step: 230, loss: 0.0031408530194312334

```

```

step: 240, loss: 6.78477153996937e-05
step: 250, loss: 2.495868830010295e-05
step: 260, loss: 0.0005508774193003774
step: 270, loss: 5.253779818303883e-05
step: 280, loss: 0.005932077765464783
step: 290, loss: 0.00025642532273195684
step: 300, loss: 0.0006174316513352096
step: 310, loss: 4.243486910127103e-05
step: 320, loss: 2.690702604013495e-05
step: 330, loss: 2.18298337131273e-05
step: 340, loss: 6.965645297896117e-05
step: 350, loss: 0.021644731983542442
step: 360, loss: 0.0009655180620029569
step: 370, loss: 8.841167436912656e-05
step: 380, loss: 3.256494164816104e-05
step: 390, loss: 2.3923283151816577e-05
step: 400, loss: 2.2813203031546436e-05
step: 410, loss: 8.183595491573215e-05
step: 420, loss: 0.001402902533300221
step: 430, loss: 6.506594945676625e-05
step: 440, loss: 0.002616873709484935
step: 450, loss: 0.0027796127833426
step: 460, loss: 8.034914208110422e-05
step: 470, loss: 6.653463788097724e-05
step: 480, loss: 3.796276723733172e-05
step: 490, loss: 0.020948518067598343
step: 500, loss: 0.0022870583925396204
step: 510, loss: 0.00011312645801808685
step: 520, loss: 0.018244989216327667
step: 530, loss: 8.915010403143242e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9332087809434843, f1=0.9237918215613383,
best_f1=0.9265799256505577
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00012437551049515605
step: 10, loss: 4.4262069422984496e-05
step: 20, loss: 0.02569551020860672
step: 30, loss: 0.00184530857950449
step: 40, loss: 2.1844742150278762e-05
step: 50, loss: 2.411681634839624e-05
step: 60, loss: 2.0924557247781195e-05
step: 70, loss: 1.8410017219139263e-05
step: 80, loss: 2.1017664039391093e-05
step: 90, loss: 4.059476123075001e-05

```

```
step: 100, loss: 1.8868237020797096e-05
step: 110, loss: 4.1165054426528513e-05
step: 120, loss: 6.542326445924118e-05
step: 130, loss: 0.0026180758140981197
step: 140, loss: 2.7022213544114493e-05
step: 150, loss: 6.19062629994005e-05
step: 160, loss: 2.1561572793871164e-05
step: 170, loss: 0.0008107625180855393
step: 180, loss: 2.4332583052455448e-05
step: 190, loss: 3.1592669984092936e-05
step: 200, loss: 2.4422390197287314e-05
step: 210, loss: 2.469762330292724e-05
step: 220, loss: 1.911043727886863e-05
step: 230, loss: 1.722544948279392e-05
step: 240, loss: 4.553321923594922e-05
step: 250, loss: 0.06513293087482452
step: 260, loss: 4.328857539803721e-05
step: 270, loss: 3.8456240872619674e-05
step: 280, loss: 4.553413964458741e-05
step: 290, loss: 2.3345772206084803e-05
step: 300, loss: 3.1444127671420574e-05
step: 310, loss: 0.00013371778186410666
step: 320, loss: 2.449688508932013e-05
step: 330, loss: 0.0013409246457740664
step: 340, loss: 2.2403502953238785e-05
step: 350, loss: 0.06900156289339066
step: 360, loss: 0.00034304324071854353
step: 370, loss: 6.613926962018013e-05
step: 380, loss: 0.000233284619753249
step: 390, loss: 0.000378457858460024
step: 400, loss: 3.133112477371469e-05
step: 410, loss: 2.6858351702685468e-05
step: 420, loss: 4.035466554341838e-05
step: 430, loss: 3.9812617615098134e-05
step: 440, loss: 3.248664870625362e-05
step: 450, loss: 0.016630899161100388
step: 460, loss: 1.8018905393546447e-05
step: 470, loss: 0.019933801144361496
step: 480, loss: 0.007521701976656914
step: 490, loss: 3.84649247280322e-05
step: 500, loss: 6.61946542095393e-05
step: 510, loss: 2.4753950128797442e-05
step: 520, loss: 5.9941397921647877e-05
step: 530, loss: 3.905735138687305e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.934916864608076, f1=0.9258384506376948,
```

```
best_f1=0.9265799256505577
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.1487997805234045e-05
step: 10, loss: 3.61064521712251e-05
step: 20, loss: 1.8905489923781715e-05
step: 30, loss: 3.2960368116619065e-05
step: 40, loss: 0.0008287220261991024
step: 50, loss: 0.00012175441952422261
step: 60, loss: 5.408040669863112e-05
step: 70, loss: 1.74601827893639e-05
step: 80, loss: 3.9970818761503324e-05
step: 90, loss: 1.539645563752856e-05
step: 100, loss: 3.018875577254221e-05
step: 110, loss: 2.9710467060795054e-05
step: 120, loss: 3.849466884275898e-05
step: 130, loss: 0.0004477773327380419
step: 140, loss: 7.793031545588747e-05
step: 150, loss: 3.5856766771757975e-05
step: 160, loss: 3.388695404282771e-05
step: 170, loss: 0.00031386251794174314
step: 180, loss: 3.87688196497038e-05
step: 190, loss: 0.00010640732216415927
step: 200, loss: 0.00011773770529543981
step: 210, loss: 7.283878221642226e-05
step: 220, loss: 0.0005332818836905062
step: 230, loss: 5.112612052471377e-05
step: 240, loss: 0.0005883910343982279
step: 250, loss: 4.788060687133111e-05
step: 260, loss: 1.805995634640567e-05
step: 270, loss: 0.00037746515590697527
step: 280, loss: 2.623169348225929e-05
step: 290, loss: 3.0194669307093136e-05
step: 300, loss: 6.725896673742682e-05
step: 310, loss: 0.016829373314976692
step: 320, loss: 0.00012275626068003476
step: 330, loss: 1.8264816390001215e-05
step: 340, loss: 2.294364821864292e-05
step: 350, loss: 4.942290979670361e-05
step: 360, loss: 2.1211391867836937e-05
step: 370, loss: 7.01661774655804e-05
step: 380, loss: 4.280362554709427e-05
step: 390, loss: 4.551154779619537e-05
step: 400, loss: 3.7253510527079925e-05
step: 410, loss: 6.769088213331997e-05
step: 420, loss: 5.9229198086541146e-05
step: 430, loss: 2.3386541215586476e-05
```

```

step: 440, loss: 1.2744110790663399e-05
step: 450, loss: 6.373945507220924e-05
step: 460, loss: 0.00010972528980346397
step: 470, loss: 1.981049354071729e-05
step: 480, loss: 1.6852958651725203e-05
step: 490, loss: 5.271062764222734e-05
step: 500, loss: 0.00036970843211747706
step: 510, loss: 3.423774614930153e-05
step: 520, loss: 2.8247641239431687e-05
step: 530, loss: 0.0026799116749316454
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9344030202925908, f1=0.9287054409005628,
best_f1=0.9265799256505577
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 2.8874936106149107e-05
step: 10, loss: 2.2924825316295028e-05
step: 20, loss: 2.8131280487286858e-05
step: 30, loss: 3.143617868772708e-05
step: 40, loss: 2.933505857072305e-05
step: 50, loss: 0.000387810287065804
step: 60, loss: 2.0748595488839783e-05
step: 70, loss: 0.00042617725557647645
step: 80, loss: 2.4582541300333105e-05
step: 90, loss: 1.4725908840773627e-05
step: 100, loss: 4.233254730934277e-05
step: 110, loss: 0.016039066016674042
step: 120, loss: 4.754665860673413e-05
step: 130, loss: 1.4141018255031668e-05
step: 140, loss: 1.4614145584346261e-05
step: 150, loss: 0.00042079028207808733
step: 160, loss: 3.264568658778444e-05
step: 170, loss: 6.719403609167784e-05
step: 180, loss: 0.005179968662559986
step: 190, loss: 0.00013558163482230157
step: 200, loss: 2.407582178420853e-05
step: 210, loss: 0.01065079029649496
step: 220, loss: 1.2341784895397723e-05
step: 230, loss: 2.796051194309257e-05
step: 240, loss: 3.622118674684316e-05
step: 250, loss: 2.609477269288618e-05
step: 260, loss: 2.119251985277515e-05
step: 270, loss: 5.6401298934360966e-05
step: 280, loss: 4.001348133897409e-05
step: 290, loss: 3.997904059360735e-05

```

```

step: 300, loss: 0.0005621368181891739
step: 310, loss: 0.00010085816757055
step: 320, loss: 3.057217691093683e-05
step: 330, loss: 2.192276770074386e-05
step: 340, loss: 0.0001328147918684408
step: 350, loss: 4.695071038440801e-05
step: 360, loss: 2.2380931113730185e-05
step: 370, loss: 9.157893509836867e-05
step: 380, loss: 1.5884301319601946e-05
step: 390, loss: 1.6592184692854062e-05
step: 400, loss: 1.5403813449665904e-05
step: 410, loss: 4.3802938307635486e-05
step: 420, loss: 0.00033310253638774157
step: 430, loss: 2.0350864360807464e-05
step: 440, loss: 2.715663686103653e-05
step: 450, loss: 0.00021299117361195385
step: 460, loss: 1.5556608559563756e-05
step: 470, loss: 9.084848716156557e-05
step: 480, loss: 2.1237066903267987e-05
step: 490, loss: 3.9260517951333895e-05
step: 500, loss: 0.004853128921240568
step: 510, loss: 0.00012619754124898463
step: 520, loss: 0.00010139083315152675
step: 530, loss: 0.00010256053792545572
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9363507779349364, f1=0.9276995305164318,
best_f1=0.9265799256505577
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True

```



```

master_weights      : True
loss_scale          : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
5742it [00:14, 395.85it/s]
load_f1 = 0.9355140186915888
real_f1 = 0.9345099860659545
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:11, 380.61it/s]

```

1.1.4 Amazon-Google - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
    --task Structured/Amazon-Google \
    --batch_size 32 \
    --max_len 256 \
    --lr 3e-5 \
    --n_epochs 15 \
    --finetuning \
    --lm distilbert \
    --fp16 \
    --save_model \
    --da entry_swap \
    --da span_del \
    --summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Amazon-Google.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
    --task Structured/Amazon-Google \
    --input_path input/input_small.jsonl \

```

```

--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:

UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order:

`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

```

"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)

```

```
step: 0, loss: 0.8458989858627319
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0

```
step: 10, loss: 0.06082675978541374
```

```
step: 20, loss: 0.38093215227127075
```

```
step: 30, loss: 0.37258338928222656
```

```

step: 40, loss: 0.4954724907875061
step: 50, loss: 0.2887062132358551
step: 60, loss: 0.3615703582763672
step: 70, loss: 0.1975320726633072
step: 80, loss: 0.32019171118736267
step: 90, loss: 0.3635731339454651
step: 100, loss: 0.09463983029127121
step: 110, loss: 0.28802046179771423
step: 120, loss: 0.2412312626838684
step: 130, loss: 0.2185136377811432
step: 140, loss: 0.21413226425647736
step: 150, loss: 0.3522283136844635
step: 160, loss: 0.24053886532783508
step: 170, loss: 0.09303493052721024
step: 180, loss: 0.17835237085819244
step: 190, loss: 0.17234835028648376
step: 200, loss: 0.12402842193841934
step: 210, loss: 0.3260080814361572
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.6336996336996338, f1=0.656880733944954,
best_f1=0.656880733944954
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06703605502843857
step: 10, loss: 0.0648796483874321
step: 20, loss: 0.23301878571510315
step: 30, loss: 0.14291022717952728
step: 40, loss: 0.07948514819145203
step: 50, loss: 0.12591256201267242
step: 60, loss: 0.05733329430222511
step: 70, loss: 0.1746024638414383
step: 80, loss: 0.15795722603797913
step: 90, loss: 0.1301608681678772
step: 100, loss: 0.07408404350280762
step: 110, loss: 0.12252878397703171
step: 120, loss: 0.14456333220005035
step: 130, loss: 0.16265149414539337
step: 140, loss: 0.11976093798875809
step: 150, loss: 0.25617820024490356
step: 160, loss: 0.11212874203920364
step: 170, loss: 0.18075650930404663

```

```

step: 180, loss: 0.15600967407226562
step: 190, loss: 0.15464839339256287
step: 200, loss: 0.3210219442844391
step: 210, loss: 0.08089347183704376
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.6575875486381323, f1=0.68762278978389,
best_f1=0.68762278978389
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.12420695275068283
step: 10, loss: 0.16398565471172333
step: 20, loss: 0.21583187580108643
step: 30, loss: 0.1140584722161293
step: 40, loss: 0.293841153383255
step: 50, loss: 0.1404622346162796
step: 60, loss: 0.308960497379303
step: 70, loss: 0.1151924580335617
step: 80, loss: 0.03158968314528465
step: 90, loss: 0.12101996690034866
step: 100, loss: 0.027227727696299553
step: 110, loss: 0.36890357732772827
step: 120, loss: 0.17943117022514343
step: 130, loss: 0.08216111361980438
step: 140, loss: 0.2713148295879364
step: 150, loss: 0.15023136138916016
step: 160, loss: 0.0991792157292366
step: 170, loss: 0.10208847373723984
step: 180, loss: 0.10612470656633377
step: 190, loss: 0.1948557049036026
step: 200, loss: 0.15953120589256287
step: 210, loss: 0.15023009479045868
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.6791666666666666, f1=0.6979591836734693,
best_f1=0.6979591836734693
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06392371654510498
step: 10, loss: 0.027336586266756058
step: 20, loss: 0.11796313524246216
step: 30, loss: 0.12823162972927094
step: 40, loss: 0.08762390166521072
step: 50, loss: 0.041344840079545975
step: 60, loss: 0.053927961736917496
step: 70, loss: 0.10682909190654755
step: 80, loss: 0.17336732149124146
step: 90, loss: 0.04120228812098503
step: 100, loss: 0.1579994112253189
step: 110, loss: 0.23413702845573425
step: 120, loss: 0.07273685932159424
step: 130, loss: 0.16464123129844666
step: 140, loss: 0.17995508015155792
step: 150, loss: 0.1280316263437271
step: 160, loss: 0.03631991147994995
step: 170, loss: 0.03322410210967064
step: 180, loss: 0.19280315935611725
step: 190, loss: 0.06406673789024353
step: 200, loss: 0.12173113226890564
step: 210, loss: 0.07217570394277573
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.6749555950266429, f1=0.6916221033868093,
best_f1=0.6979591836734693
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.040610406547784805
step: 10, loss: 0.14683593809604645
step: 20, loss: 0.06127643585205078
step: 30, loss: 0.09716561436653137
step: 40, loss: 0.07424571365118027
step: 50, loss: 0.19783124327659607
step: 60, loss: 0.10136819630861282
step: 70, loss: 0.20280230045318604
step: 80, loss: 0.03482868894934654
step: 90, loss: 0.046465203166007996
step: 100, loss: 0.054782453924417496
step: 110, loss: 0.018932342529296875
step: 120, loss: 0.015197601169347763
step: 130, loss: 0.049826234579086304
step: 140, loss: 0.09485234320163727
step: 150, loss: 0.026774490252137184

```

```

step: 160, loss: 0.013953388668596745
step: 170, loss: 0.0244764257222414
step: 180, loss: 0.11097732186317444
step: 190, loss: 0.09969700872898102
step: 200, loss: 0.07095740735530853
step: 210, loss: 0.018914662301540375
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.6925925925925926, f1=0.6862385321100918,
best_f1=0.6862385321100918
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09771854430437088
step: 10, loss: 0.013773923739790916
step: 20, loss: 0.03610658273100853
step: 30, loss: 0.11311031132936478
step: 40, loss: 0.006411868613213301
step: 50, loss: 0.04719875380396843
step: 60, loss: 0.07466335594654083
step: 70, loss: 0.002311748918145895
step: 80, loss: 0.05631944537162781
step: 90, loss: 0.1229219064116478
step: 100, loss: 0.05422564223408699
step: 110, loss: 0.08620637655258179
step: 120, loss: 0.01393401063978672
step: 130, loss: 0.12240926921367645
step: 140, loss: 0.05291050672531128
step: 150, loss: 0.01981196366250515
step: 160, loss: 0.10282883793115616
step: 170, loss: 0.04536950960755348
step: 180, loss: 0.0075058648362755775
step: 190, loss: 0.07286808639764786
step: 200, loss: 0.025327064096927643
step: 210, loss: 0.06069977954030037
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.6812386156648452, f1=0.6861313868613138,
best_f1=0.6862385321100918
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.053495198488235474
step: 10, loss: 0.11323821544647217
step: 20, loss: 0.01388476975262165
step: 30, loss: 0.056113749742507935
step: 40, loss: 0.01113598607480526
step: 50, loss: 0.02886631153523922
step: 60, loss: 0.17931802570819855
step: 70, loss: 0.010571752674877644
step: 80, loss: 0.04768966883420944
step: 90, loss: 0.007087579928338528
step: 100, loss: 0.017810121178627014
step: 110, loss: 0.10645996034145355
step: 120, loss: 0.11686910688877106
step: 130, loss: 0.025520889088511467
step: 140, loss: 0.0043150633573532104
step: 150, loss: 0.05025356635451317
step: 160, loss: 0.011529544368386269
step: 170, loss: 0.0353982150554657
step: 180, loss: 0.0008545409655198455
step: 190, loss: 0.014075910672545433
step: 200, loss: 0.08345825970172882
step: 210, loss: 0.011238222010433674
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.6743295019157088, f1=0.6666666666666666,
best_f1=0.6862385321100918
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03995644301176071
step: 10, loss: 0.031705617904663086
step: 20, loss: 0.09468680620193481
step: 30, loss: 0.1590450257062912
step: 40, loss: 0.032056763768196106
step: 50, loss: 0.06375005841255188
step: 60, loss: 0.18149159848690033
step: 70, loss: 0.0112153897061944
step: 80, loss: 0.01489940658211708
step: 90, loss: 0.03581082075834274
step: 100, loss: 0.009492160752415657
step: 110, loss: 0.009186566807329655
step: 120, loss: 0.02585294470191002
step: 130, loss: 0.08564171940088272
step: 140, loss: 0.0010307180928066373
step: 150, loss: 0.07665912806987762
step: 160, loss: 0.146675705909729
step: 170, loss: 0.18868103623390198

```

```

step: 180, loss: 0.028926612809300423
step: 190, loss: 0.12732410430908203
step: 200, loss: 0.04308844730257988
step: 210, loss: 0.09237608313560486
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.6642984014209592, f1=0.673721340388007,
best_f1=0.6862385321100918
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00496737752109766
step: 10, loss: 0.07069768011569977
step: 20, loss: 0.06922022253274918
step: 30, loss: 0.06096421182155609
step: 40, loss: 0.13503679633140564
step: 50, loss: 0.02116112783551216
step: 60, loss: 0.05976005643606186
step: 70, loss: 0.03641822934150696
step: 80, loss: 0.014020606875419617
step: 90, loss: 0.03689500316977501
step: 100, loss: 0.04561387747526169
step: 110, loss: 0.009170064702630043
step: 120, loss: 0.011818073689937592
step: 130, loss: 0.00442567327991128
step: 140, loss: 0.011947301216423512
step: 150, loss: 0.01109904982149601
step: 160, loss: 0.003081294009461999
step: 170, loss: 0.341856986284256
step: 180, loss: 0.013263272121548653
step: 190, loss: 0.06555923074483871
step: 200, loss: 0.02624446526169777
step: 210, loss: 0.12667986750602722
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.6848249027237354, f1=0.66796875, best_f1=0.6862385321100918
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.023812806233763695
step: 10, loss: 0.04763977602124214
step: 20, loss: 0.03159768879413605
step: 30, loss: 0.016123782843351364
step: 40, loss: 0.15208907425403595
step: 50, loss: 0.01297035999596119
step: 60, loss: 0.0012702863896265626

```



```

step: 70, loss: 0.0481758713722229
step: 80, loss: 0.0004476614121813327
step: 90, loss: 0.025531671941280365
step: 100, loss: 0.006204158999025822
step: 110, loss: 0.0017633171519264579
step: 120, loss: 0.0015856333775445819
step: 130, loss: 0.0029832483269274235
step: 140, loss: 0.0015261084772646427
step: 150, loss: 0.11186258494853973
step: 160, loss: 0.1617741882801056
step: 170, loss: 0.11060070991516113
step: 180, loss: 0.022589359432458878
step: 190, loss: 0.04578380659222603
step: 200, loss: 0.12429690361022949
step: 210, loss: 0.2183726727962494
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.6842105263157895, f1=0.6802973977695167,
best_f1=0.6862385321100918
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01654260978102684
step: 10, loss: 0.011046705767512321
step: 20, loss: 0.004512476734817028
step: 30, loss: 0.03254779055714607
step: 40, loss: 0.03104567714035511
step: 50, loss: 0.006520126946270466
step: 60, loss: 0.08584403246641159
step: 70, loss: 0.0028268597088754177
step: 80, loss: 0.12134962528944016
step: 90, loss: 0.13525761663913727
step: 100, loss: 0.1066126599907875
step: 110, loss: 0.008119593374431133
step: 120, loss: 0.0028351403307169676
step: 130, loss: 0.027439827099442482
step: 140, loss: 0.0065791429951786995
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 150, loss: 0.09304141253232956
step: 160, loss: 0.020495139062404633
step: 170, loss: 0.062268298119306564
step: 180, loss: 0.04051555320620537
step: 190, loss: 0.021616771817207336
step: 200, loss: 0.000343776758801192
step: 210, loss: 0.01667034812271595
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 11: dev_f1=0.6956521739130436, f1=0.6850393700787402,
best_f1=0.6850393700787402
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007573016337119043
step: 10, loss: 0.014891250059008598
step: 20, loss: 0.005011529661715031
step: 30, loss: 0.005518130026757717
step: 40, loss: 0.1505596786737442
step: 50, loss: 0.032199498265981674
step: 60, loss: 0.004859384149312973
step: 70, loss: 0.0031632145401090384
step: 80, loss: 0.009390326216816902
step: 90, loss: 0.002824767492711544
step: 100, loss: 0.0006383597501553595
step: 110, loss: 0.015898151323199272
step: 120, loss: 0.09592165052890778
step: 130, loss: 0.017603004351258278
step: 140, loss: 0.01789654791355133
step: 150, loss: 0.003963501658290625
step: 160, loss: 0.022602565586566925
step: 170, loss: 0.004039403051137924
step: 180, loss: 0.0013618555385619402
step: 190, loss: 0.11389338225126266
step: 200, loss: 0.05175041779875755
step: 210, loss: 0.009098341688513756
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.6823529411764706, f1=0.683111954459203,
best_f1=0.6850393700787402
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0012736925855278969
step: 10, loss: 0.006890930235385895
step: 20, loss: 0.0027923339512199163
step: 30, loss: 0.022087596356868744
step: 40, loss: 0.0006228964775800705
step: 50, loss: 0.06409784406423569
step: 60, loss: 0.03346305713057518
step: 70, loss: 0.11516232043504715

```

```

step: 80, loss: 0.18379519879817963
step: 90, loss: 0.06368566304445267
step: 100, loss: 0.0689605325460434
step: 110, loss: 0.007617400959134102
step: 120, loss: 0.017157748341560364
step: 130, loss: 0.004398055374622345
step: 140, loss: 0.006609059404581785
step: 150, loss: 0.0014940255787223577
step: 160, loss: 0.04852694272994995
step: 170, loss: 0.07235506922006607
step: 180, loss: 0.21128207445144653
step: 190, loss: 0.005593674723058939
step: 200, loss: 0.007286841049790382
step: 210, loss: 0.011984409764409065
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.6842105263157895, f1=0.6867469879518072,
best_f1=0.6850393700787402
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005230996757745743
step: 10, loss: 0.0035959105007350445
step: 20, loss: 0.0016679682303220034
step: 30, loss: 0.011477479711174965
step: 40, loss: 0.0020603311713784933
step: 50, loss: 0.02532687596976757
step: 60, loss: 0.02419566549360752
step: 70, loss: 0.00744572514668107
step: 80, loss: 0.05909115821123123
step: 90, loss: 0.1262306571006775
step: 100, loss: 0.0025904683861881495
step: 110, loss: 0.12520353496074677
step: 120, loss: 0.010126658715307713
step: 130, loss: 0.0018226063111796975
step: 140, loss: 0.018912523984909058
step: 150, loss: 0.04171071946620941
step: 160, loss: 0.004495818633586168
step: 170, loss: 0.004407590255141258
step: 180, loss: 0.00535175483673811
step: 190, loss: 0.004928923211991787
step: 200, loss: 0.0017344681546092033
step: 210, loss: 0.0022779235150665045
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.6826347305389221, f1=0.6942800788954636,

```

```

best_f1=0.6850393700787402
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.021073095500469208
step: 10, loss: 0.009065195918083191
step: 20, loss: 0.0007274650852195919
step: 30, loss: 0.0009691692539490759
step: 40, loss: 0.0011660474119707942
step: 50, loss: 0.002277195453643799
step: 60, loss: 0.026905842125415802
step: 70, loss: 0.0009802323766052723
step: 80, loss: 0.0028630252927541733
step: 90, loss: 0.001924209762364626
step: 100, loss: 0.0073852138593792915
step: 110, loss: 0.00056800915626809
step: 120, loss: 0.07701165229082108
step: 130, loss: 0.053623344749212265
step: 140, loss: 0.0014414989855140448
step: 150, loss: 0.003093754407018423
step: 160, loss: 0.01923973485827446
step: 170, loss: 0.05446970462799072
step: 180, loss: 0.008963357657194138
step: 190, loss: 0.032224372029304504
step: 200, loss: 0.001059960457496345
step: 210, loss: 0.06826227903366089
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.6805845511482255, f1=0.6763485477178423,
best_f1=0.6850393700787402
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16

```

```

patch_torch_functions : False
keep_batchnorm_fp32   : True
master_weights        : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2293it [00:03, 576.12it/s]
load_f1 = 0.686046511627907
real_f1 = 0.683495145631068
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:11, 374.26it/s]

```

1.1.5 Walmart-Amazon - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
    --task Structured/Walmart-Amazon \
    --batch_size 32 \
    --max_len 256 \
    --lr 3e-5 \
    --n_epochs 15 \
    --finetuning \
    --lm distilbert \
    --fp16 \
    --save_model \
    --da entry_swap \
    --da span_del \
    --summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Walmart-Amazon.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \

```

```

--task Structured/Walmart-Amazon \
--input_path input/input_small.jsonl \
--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/

```

Selected optimization level O2: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled          : True
opt_level        : O2
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled          : True
opt_level        : O2
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

warnings.warn("An input tensor was not cuda.")

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:

UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order: `optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)

step: 0, loss: 0.863209068775177

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.1610778421163559

```

step: 20, loss: 0.15920132398605347
step: 30, loss: 0.5059854984283447
step: 40, loss: 0.2583993077278137
step: 50, loss: 0.30946463346481323
step: 60, loss: 0.36182641983032227
step: 70, loss: 0.17257684469223022
step: 80, loss: 0.5050814151763916
step: 90, loss: 0.2319679707288742
step: 100, loss: 0.21787285804748535
step: 110, loss: 0.23836782574653625
step: 120, loss: 0.4141973555088043
step: 130, loss: 0.34651780128479004
step: 140, loss: 0.3232775330543518
step: 150, loss: 0.28146666288375854
step: 160, loss: 0.17417733371257782
step: 170, loss: 0.3676205575466156
step: 180, loss: 0.2406197190284729
step: 190, loss: 0.09249872714281082
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.610079575596817, f1=0.6267029972752044,
best_f1=0.6267029972752044
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.20065782964229584
step: 10, loss: 0.03598771244287491
step: 20, loss: 0.1317971795797348
step: 30, loss: 0.10889367014169693
step: 40, loss: 0.4169328808784485
step: 50, loss: 0.28768685460090637
step: 60, loss: 0.06255556643009186
step: 70, loss: 0.1299024075269699
step: 80, loss: 0.13948403298854828
step: 90, loss: 0.13306574523448944
step: 100, loss: 0.27385252714157104
step: 110, loss: 0.031061455607414246
step: 120, loss: 0.12267003953456879
step: 130, loss: 0.13276629149913788
step: 140, loss: 0.11247358471155167
step: 150, loss: 0.10342489182949066
step: 160, loss: 0.07078856974840164
step: 170, loss: 0.09467020630836487

```

```

step: 180, loss: 0.12409452348947525
step: 190, loss: 0.10346963256597519
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.7538461538461538, f1=0.792079207920792,
best_f1=0.792079207920792
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.16352394223213196
step: 10, loss: 0.18359088897705078
step: 20, loss: 0.08030321449041367
step: 30, loss: 0.019595561549067497
step: 40, loss: 0.021036813035607338
step: 50, loss: 0.17514878511428833
step: 60, loss: 0.04804180562496185
step: 70, loss: 0.14313313364982605
step: 80, loss: 0.15232476592063904
step: 90, loss: 0.07852771133184433
step: 100, loss: 0.058448243886232376
step: 110, loss: 0.17027370631694794
step: 120, loss: 0.010563407093286514
step: 130, loss: 0.03726803883910179
step: 140, loss: 0.02964167110621929
step: 150, loss: 0.06983184814453125
step: 160, loss: 0.10890884697437286
step: 170, loss: 0.051655445247888565
step: 180, loss: 0.07874187082052231
step: 190, loss: 0.08228418231010437
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.7842105263157894, f1=0.7956989247311829,
best_f1=0.7956989247311829
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08028478920459747
step: 10, loss: 0.07303066551685333

```



```

step: 20, loss: 0.028693677857518196
step: 30, loss: 0.032959066331386566
step: 40, loss: 0.03947879374027252
step: 50, loss: 0.027696901932358742
step: 60, loss: 0.14739303290843964
step: 70, loss: 0.03741493821144104
step: 80, loss: 0.009085779078304768
step: 90, loss: 0.014397026039659977
step: 100, loss: 0.022378837689757347
step: 110, loss: 0.04147405922412872
step: 120, loss: 0.03269697725772858
step: 130, loss: 0.14550404250621796
step: 140, loss: 0.027273831889033318
step: 150, loss: 0.044050537049770355
step: 160, loss: 0.08860265463590622
step: 170, loss: 0.016313403844833374
step: 180, loss: 0.08740219473838806
step: 190, loss: 0.039283428341150284
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.7979274611398963, f1=0.7821522309711286,
best_f1=0.7821522309711286
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08089067041873932
step: 10, loss: 0.012048422358930111
step: 20, loss: 0.040260523557662964
step: 30, loss: 0.02930201031267643
step: 40, loss: 0.021026942878961563
step: 50, loss: 0.013122213073074818
step: 60, loss: 0.0058707608841359615
step: 70, loss: 0.0021070013754069805
step: 80, loss: 0.005598913878202438
step: 90, loss: 0.030301308259367943
step: 100, loss: 0.04985542967915535
step: 110, loss: 0.005688996519893408
step: 120, loss: 0.004799882415682077
step: 130, loss: 0.008611241355538368
step: 140, loss: 0.015095565468072891
step: 150, loss: 0.014505762606859207
step: 160, loss: 0.015001792460680008
step: 170, loss: 0.04008841514587402

```

```

step: 180, loss: 0.06823202222585678
step: 190, loss: 0.11486285924911499
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.7766497461928934, f1=0.7589743589743588,
best_f1=0.7821522309711286
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04756713658571243
step: 10, loss: 0.001815652591176331
step: 20, loss: 0.0021257493644952774
step: 30, loss: 0.0019958200864493847
step: 40, loss: 0.1083936095237732
step: 50, loss: 0.010756701231002808
step: 60, loss: 0.0016886427765712142
step: 70, loss: 0.07720420509576797
step: 80, loss: 0.11211169511079788
step: 90, loss: 0.028729498386383057
step: 100, loss: 0.00507945055142045
step: 110, loss: 0.00966658815741539
step: 120, loss: 0.012239227071404457
step: 130, loss: 0.01371287927031517
step: 140, loss: 0.008814403787255287
step: 150, loss: 0.0066105336882174015
step: 160, loss: 0.012641489505767822
step: 170, loss: 0.011666578240692616
step: 180, loss: 0.0818423181772232
step: 190, loss: 0.01629399135708809
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.761904761904762, f1=0.7883597883597884,
best_f1=0.7821522309711286
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014575914479792118
step: 10, loss: 0.008842327632009983
step: 20, loss: 0.0009357251110486686
step: 30, loss: 0.03261198475956917
step: 40, loss: 0.053794048726558685
step: 50, loss: 0.06261458992958069
step: 60, loss: 0.010290281847119331
step: 70, loss: 0.0022345934994518757
step: 80, loss: 0.011964102275669575
step: 90, loss: 0.0066167027689516544

```

```

step: 100, loss: 0.002260896610096097
step: 110, loss: 0.012789719738066196
step: 120, loss: 0.01185722928494215
step: 130, loss: 0.008256426081061363
step: 140, loss: 0.00348414434120059
step: 150, loss: 0.028510164469480515
step: 160, loss: 0.00131339265499264
step: 170, loss: 0.010256476700305939
step: 180, loss: 0.009208970703184605
step: 190, loss: 0.011305442079901695
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.7937336814621411, f1=0.7956403269754768,
best_f1=0.7821522309711286
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007503096014261246
step: 10, loss: 0.02583271451294422
step: 20, loss: 0.012591823004186153
step: 30, loss: 0.00315472693182528
step: 40, loss: 0.02908601053059101
step: 50, loss: 0.0035295062698423862
step: 60, loss: 0.0017566293245181441
step: 70, loss: 0.003300765762105584
step: 80, loss: 0.047097235918045044
step: 90, loss: 0.012173941358923912
step: 100, loss: 0.052048854529857635
step: 110, loss: 0.015546959824860096
step: 120, loss: 0.04273798689246178
step: 130, loss: 0.00227869744412601
step: 140, loss: 0.019921213388442993
step: 150, loss: 0.033596016466617584
step: 160, loss: 0.0011641657911241055
step: 170, loss: 0.011057074181735516
step: 180, loss: 0.012489048764109612
step: 190, loss: 0.15143917500972748
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 8: dev_f1=0.8062827225130891, f1=0.8041775456919059,
best_f1=0.8041775456919059
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.005464053712785244
step: 10, loss: 0.0008625291520729661
step: 20, loss: 0.025309549644589424
step: 30, loss: 0.0007331466185860336
step: 40, loss: 0.0022696482483297586
step: 50, loss: 0.010773208923637867
step: 60, loss: 0.0030888174660503864
step: 70, loss: 0.07550415396690369
step: 80, loss: 0.0047427210956811905
step: 90, loss: 0.007391318213194609
step: 100, loss: 0.002657939214259386
step: 110, loss: 0.04462408646941185
step: 120, loss: 0.07284153252840042
step: 130, loss: 0.0011643834877759218
step: 140, loss: 0.004059332888573408
step: 150, loss: 0.013400630094110966
step: 160, loss: 0.0010989679722115397
step: 170, loss: 0.011224021203815937
step: 180, loss: 0.0022712484933435917
step: 190, loss: 0.007365938741713762
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

```
epoch 9: dev_f1=0.7886597938144331, f1=0.7968337730870713,
best_f1=0.8041775456919059
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.00037158330087549984
step: 10, loss: 0.00046833668602630496
step: 20, loss: 0.0004714771930593997
step: 30, loss: 0.0004745639453176409
step: 40, loss: 0.001533239963464439
step: 50, loss: 0.0023660368751734495
step: 60, loss: 0.015233641490340233
step: 70, loss: 0.003423691261559725
step: 80, loss: 0.03669872507452965
step: 90, loss: 0.002617324236780405
step: 100, loss: 0.0029825458768755198
step: 110, loss: 0.013646858744323254
step: 120, loss: 0.025822386145591736
step: 130, loss: 0.005024643149226904
step: 140, loss: 0.0016562271630391479
step: 150, loss: 0.0013575954362750053
step: 160, loss: 0.0017428818391636014
step: 170, loss: 0.002042958978563547
```

```

step: 180, loss: 0.004550440236926079
step: 190, loss: 0.003996419254690409
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.7945205479452055, f1=0.7955801104972376,
best_f1=0.8041775456919059
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006348499562591314
step: 10, loss: 0.0016178160440176725
step: 20, loss: 0.030075255781412125
step: 30, loss: 0.004924983251839876
step: 40, loss: 0.0009746336145326495
step: 50, loss: 0.0058752140030264854
step: 60, loss: 0.0009454251267015934
step: 70, loss: 0.00037324291770346463
step: 80, loss: 0.015545709989964962
step: 90, loss: 0.00463476637378335
step: 100, loss: 0.0025390544906258583
step: 110, loss: 0.0007535315817221999
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 120, loss: 0.0048825060948729515
step: 130, loss: 0.002085393527522683
step: 140, loss: 0.0005825693951919675
step: 150, loss: 0.004063216503709555
step: 160, loss: 0.0007107086130417883
step: 170, loss: 0.0018581411568447948
step: 180, loss: 0.16388796269893646
step: 190, loss: 0.004149965941905975
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.8, f1=0.8067226890756303, best_f1=0.8041775456919059
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012333236634731293
step: 10, loss: 0.007097323890775442
step: 20, loss: 0.12571415305137634
step: 30, loss: 0.0024482731241732836
step: 40, loss: 0.0021316315978765488
step: 50, loss: 0.009550163522362709
step: 60, loss: 0.002958557801321149
step: 70, loss: 0.0028643396217375994
step: 80, loss: 0.0004233882937114686
step: 90, loss: 0.0010744418250396848

```

```

step: 100, loss: 0.002848502481356263
step: 110, loss: 0.0005805889959447086
step: 120, loss: 0.0005957674002274871
step: 130, loss: 0.00043526702211238444
step: 140, loss: 0.001187443034723401
step: 150, loss: 0.0019300660351291299
step: 160, loss: 0.06886270642280579
step: 170, loss: 0.01675407588481903
step: 180, loss: 0.0017414354952052236
step: 190, loss: 0.020549234002828598
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.796068796068796, f1=0.7758186397984886,
best_f1=0.8041775456919059
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.12915171682834625
step: 10, loss: 0.003011328401044011
step: 20, loss: 0.002500317059457302
step: 30, loss: 0.0008574957610107958
step: 40, loss: 0.002933507552370429
step: 50, loss: 0.16130684316158295
step: 60, loss: 0.0016710906056687236
step: 70, loss: 0.021052315831184387
step: 80, loss: 0.006382256280630827
step: 90, loss: 0.001596239977516234
step: 100, loss: 0.002579019172117114
step: 110, loss: 0.011119725182652473
step: 120, loss: 0.004895340651273727
step: 130, loss: 0.003110211342573166
step: 140, loss: 0.0010975166223943233
step: 150, loss: 0.0020597733091562986
step: 160, loss: 0.005135984625667334
step: 170, loss: 0.0012520900927484035
step: 180, loss: 0.007483061868697405
step: 190, loss: 0.0045202793553471565
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.8041237113402062, f1=0.7926509186351706,
best_f1=0.8041775456919059
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.018668370321393013
step: 10, loss: 0.0009507092763669789

```

```

step: 20, loss: 0.0005766847752965987
step: 30, loss: 0.007287611253559589
step: 40, loss: 0.00033029718906618655
step: 50, loss: 0.0022123095113784075
step: 60, loss: 0.0015607918612658978
step: 70, loss: 0.0024968483485281467
step: 80, loss: 0.0006924837944097817
step: 90, loss: 0.0007487408583983779
step: 100, loss: 0.0009715981432236731
step: 110, loss: 0.0022773563396185637
step: 120, loss: 0.0030911010690033436
step: 130, loss: 0.0015462972223758698
step: 140, loss: 0.0016532227164134383
step: 150, loss: 0.0006648791604675353
step: 160, loss: 0.0005905390135012567
step: 170, loss: 0.007119060028344393
step: 180, loss: 0.0011678861919790506
step: 190, loss: 0.0002711659180931747
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.7927461139896373, f1=0.7926509186351706,
best_f1=0.8041775456919059
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0023911858443170786
step: 10, loss: 0.0024286117404699326
step: 20, loss: 0.001093282480724156
step: 30, loss: 0.0028694889042526484
step: 40, loss: 0.004523991607129574
step: 50, loss: 0.001364808646030724
step: 60, loss: 0.008350192569196224
step: 70, loss: 0.0007039696793071926
step: 80, loss: 0.001673276536166668
step: 90, loss: 0.012589550577104092
step: 100, loss: 0.0003579509793780744
step: 110, loss: 0.00528199365362525
step: 120, loss: 0.0008699483587406576
step: 130, loss: 0.0007016416639089584
step: 140, loss: 0.008147629909217358
step: 150, loss: 0.0005301529308781028
step: 160, loss: 0.0013167470460757613
step: 170, loss: 0.0018509005894884467
step: 180, loss: 0.0017088644672185183
step: 190, loss: 0.008885368704795837
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.7928388746803069, f1=0.7885117493472584,
best_f1=0.8041775456919059
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2049it [00:07, 274.31it/s]
load_f1 = 0.6085192697768762
real_f1 = 0.5812619502868069
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:11, 368.92it/s]

```


1.2 DA TEXTUAL

1.2.1 Abt-Buy - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Textual/Abt-Buy \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model \
  --da entry_swap \
  --da span_shuffle \
  --summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Abt-Buy.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Textual/Abt-Buy \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic
```

Processing user overrides (additional kwargs that are not None)...

```

After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8559926748275757
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.2325299084186554
step: 20, loss: 0.14835302531719208
step: 30, loss: 0.23891931772232056
step: 40, loss: 0.3150731325149536
step: 50, loss: 0.3783982992172241
step: 60, loss: 0.4356561601161957
step: 70, loss: 0.31675517559051514
step: 80, loss: 0.2518249750137329
step: 90, loss: 0.39534273743629456
step: 100, loss: 0.2323702573776245
step: 110, loss: 0.17802204191684723
step: 120, loss: 0.5184791088104248
step: 130, loss: 0.3723081350326538
step: 140, loss: 0.32628780603408813
step: 150, loss: 0.14346076548099518
step: 160, loss: 0.2671665549278259
step: 170, loss: 0.2772006392478943
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or

```

```

loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.6666666666666666, f1=0.6806722689075629,
best_f1=0.6806722689075629
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2195250540971756
step: 10, loss: 0.04677249491214752
step: 20, loss: 0.28120341897010803
step: 30, loss: 0.16029448807239532
step: 40, loss: 0.13976749777793884
step: 50, loss: 0.20209765434265137
step: 60, loss: 0.049440979957580566
step: 70, loss: 0.09180672466754913
step: 80, loss: 0.17549461126327515
step: 90, loss: 0.257095068693161
step: 100, loss: 0.1214267909526825
step: 110, loss: 0.209659144282341
step: 120, loss: 0.037764277309179306
step: 130, loss: 0.024531252682209015
step: 140, loss: 0.09668617695569992
step: 150, loss: 0.11770734190940857
step: 160, loss: 0.08871898055076599
step: 170, loss: 0.0419507697224617
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.7961630695443646, f1=0.7936507936507935,
best_f1=0.7936507936507935
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05375262722373009
step: 10, loss: 0.031105773523449898
step: 20, loss: 0.06750049442052841
step: 30, loss: 0.09550528973340988
step: 40, loss: 0.03155435994267464
step: 50, loss: 0.22076551616191864
step: 60, loss: 0.12435856461524963
step: 70, loss: 0.044097237288951874
step: 80, loss: 0.11795017123222351
step: 90, loss: 0.0673016905784607
step: 100, loss: 0.03444164618849754

```

```

step: 110, loss: 0.29569512605667114
step: 120, loss: 0.015943117439746857
step: 130, loss: 0.12624700367450714
step: 140, loss: 0.004154201131314039
step: 150, loss: 0.058807965368032455
step: 160, loss: 0.16603459417819977
step: 170, loss: 0.051430121064186096
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.804177545691906, f1=0.780361757105943,
best_f1=0.780361757105943
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03742202743887901
step: 10, loss: 0.2179785519838333
step: 20, loss: 0.07511256635189056
step: 30, loss: 0.05144122615456581
step: 40, loss: 0.00824306346476078
step: 50, loss: 0.010381672531366348
step: 60, loss: 0.10073350369930267
step: 70, loss: 0.06124670431017876
step: 80, loss: 0.052440546452999115
step: 90, loss: 0.040958914905786514
step: 100, loss: 0.010898121632635593
step: 110, loss: 0.020936986431479454
step: 120, loss: 0.11897488683462143
step: 130, loss: 0.014675971120595932
step: 140, loss: 0.01897585205733776
step: 150, loss: 0.0008460463723167777
step: 160, loss: 0.09617026150226593
step: 170, loss: 0.003225575666874647
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.8106796116504854, f1=0.8037383177570093,
best_f1=0.8037383177570093
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.018305504694581032
step: 10, loss: 0.039031870663166046
step: 20, loss: 0.13089817762374878
step: 30, loss: 0.03572692349553108
step: 40, loss: 0.05755365267395973
step: 50, loss: 0.014939927496016026
step: 60, loss: 0.00856481771916151
step: 70, loss: 0.01625283993780613
step: 80, loss: 0.005375227425247431
step: 90, loss: 0.01182526908814907
step: 100, loss: 0.03172696381807327
step: 110, loss: 0.11867223680019379
step: 120, loss: 0.0280611589550972
step: 130, loss: 0.05245343595743179
step: 140, loss: 0.0944870263338089
step: 150, loss: 0.006105746142566204
step: 160, loss: 0.00867184903472662
step: 170, loss: 0.1221584603190422
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.8246913580246913, f1=0.821852731591449,
best_f1=0.821852731591449
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01873435638844967
step: 10, loss: 0.13591577112674713
step: 20, loss: 0.005547675769776106
step: 30, loss: 0.043700169771909714
step: 40, loss: 0.008895233273506165
step: 50, loss: 0.18567930161952972
step: 60, loss: 0.00838044099509716
step: 70, loss: 0.02599911577999592
step: 80, loss: 0.13848714530467987
step: 90, loss: 0.02761845290660858
step: 100, loss: 0.09742215275764465
step: 110, loss: 0.06811439245939255
step: 120, loss: 0.09049241244792938
step: 130, loss: 0.18410740792751312
step: 140, loss: 0.022747932001948357
step: 150, loss: 0.1711978018283844
step: 160, loss: 0.13094812631607056

```

```

step: 170, loss: 0.05718205124139786
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 6: dev_f1=0.83, f1=0.8126520681265207, best_f1=0.8126520681265207
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01458786427974701
step: 10, loss: 0.002870421390980482
step: 20, loss: 0.007337490096688271
step: 30, loss: 0.013695724308490753
step: 40, loss: 0.009355071932077408
step: 50, loss: 0.005722907371819019
step: 60, loss: 0.24228236079216003
step: 70, loss: 0.006854918785393238
step: 80, loss: 0.003605990670621395
step: 90, loss: 0.03809449449181557
step: 100, loss: 0.013343089260160923
step: 110, loss: 0.00030590835376642644
step: 120, loss: 0.09236259013414383
step: 130, loss: 0.053023066371679306
step: 140, loss: 0.01785777136683464
step: 150, loss: 0.04228600487112999
step: 160, loss: 0.09120442718267441
step: 170, loss: 0.0699198767542839
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.821852731591449, f1=0.7954545454545455,
best_f1=0.8126520681265207
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.022423148155212402
step: 10, loss: 0.02312697097659111
step: 20, loss: 0.012238581664860249
step: 30, loss: 0.016420649364590645
step: 40, loss: 0.00931934267282486
step: 50, loss: 0.0032898455392569304
step: 60, loss: 0.006673681084066629
step: 70, loss: 0.08867866545915604
step: 80, loss: 0.002158353105187416
step: 90, loss: 0.02151353470981121

```

```

step: 100, loss: 0.00583826145157218
step: 110, loss: 0.0441623255610466
step: 120, loss: 0.001336064306087792
step: 130, loss: 0.0010917203035205603
step: 140, loss: 0.034967076033353806
step: 150, loss: 0.0374007448554039
step: 160, loss: 0.009950762614607811
step: 170, loss: 0.0033581901807338
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.8099999999999999, f1=0.7836538461538461,
best_f1=0.8126520681265207
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007402302697300911
step: 10, loss: 0.000790534308180213
step: 20, loss: 0.00043446978088468313
step: 30, loss: 0.0456838496029377
step: 40, loss: 0.005686501041054726
step: 50, loss: 0.004233829211443663
step: 60, loss: 0.0013547358103096485
step: 70, loss: 0.024740541353821754
step: 80, loss: 0.06848090887069702
step: 90, loss: 0.0056533063761889935
step: 100, loss: 0.05700172483921051
step: 110, loss: 0.03546856343746185
step: 120, loss: 0.0019621436949819326
step: 130, loss: 0.0007180179818533361
step: 140, loss: 0.00028244321583770216
step: 150, loss: 0.0224661435931921
step: 160, loss: 0.011155432090163231
step: 170, loss: 0.01807430572807789
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.7989690721649484, f1=0.8, best_f1=0.8126520681265207
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0013499833876267076
step: 10, loss: 0.060006171464920044
step: 20, loss: 0.010522841475903988
step: 30, loss: 0.03268282115459442
step: 40, loss: 0.03968183323740959
step: 50, loss: 0.05787966400384903
step: 60, loss: 0.003896160051226616

```

```

step: 70, loss: 0.013449383899569511
step: 80, loss: 0.007076198700815439
step: 90, loss: 0.027020109817385674
step: 100, loss: 0.028095228597521782
step: 110, loss: 0.0010356188286095858
step: 120, loss: 0.020514417439699173
step: 130, loss: 0.00826639961451292
step: 140, loss: 0.011417143978178501
step: 150, loss: 0.0009833411313593388
step: 160, loss: 0.0002920048136729747
step: 170, loss: 0.004783284850418568
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.8170426065162907, f1=0.8038277511961723,
best_f1=0.8126520681265207
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0014671924291178584
step: 10, loss: 0.015411218628287315
step: 20, loss: 0.039513975381851196
step: 30, loss: 0.0011284210486337543
step: 40, loss: 0.012502776458859444
step: 50, loss: 0.08977128565311432
step: 60, loss: 0.001370111363939941
step: 70, loss: 0.0016543834935873747
step: 80, loss: 0.0012967409566044807
step: 90, loss: 0.0018332900945097208
step: 100, loss: 0.0003051420790143311
step: 110, loss: 0.006460977252572775
step: 120, loss: 0.02246806025505066
step: 130, loss: 0.0012626437237486243
step: 140, loss: 0.04107460379600525
step: 150, loss: 0.0006871569203212857
step: 160, loss: 0.013600220903754234
step: 170, loss: 0.05185398459434509
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.8010335917312662, f1=0.7980295566502462,
best_f1=0.8126520681265207
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014145773835480213
step: 10, loss: 0.003135200124233961
step: 20, loss: 0.017423057928681374

```



```

step: 30, loss: 0.002412233268842101
step: 40, loss: 0.00018401634588371962
step: 50, loss: 0.004609726835042238
step: 60, loss: 0.0008068689494393766
step: 70, loss: 0.0028695568908005953
step: 80, loss: 0.044390950351953506
step: 90, loss: 0.0005806930712424219
step: 100, loss: 0.02559615671634674
step: 110, loss: 0.00016557938943151385
step: 120, loss: 0.03560929000377655
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 130, loss: 0.0029270676895976067
step: 140, loss: 0.0007304172031581402
step: 150, loss: 0.006494682282209396
step: 160, loss: 0.06387443095445633
step: 170, loss: 0.003588548395782709
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.797843665768194, f1=0.7846153846153846,
best_f1=0.8126520681265207
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00022242152772378176
step: 10, loss: 0.05174380540847778
step: 20, loss: 0.0004156418435741216
step: 30, loss: 0.00013730343198403716
step: 40, loss: 0.0029420906212180853
step: 50, loss: 0.0334569551050663
step: 60, loss: 0.012419058009982109
step: 70, loss: 0.009311932139098644
step: 80, loss: 0.0009374815272167325
step: 90, loss: 0.00025942837237380445
step: 100, loss: 0.00011230471136514097
step: 110, loss: 0.00021743499382864684
step: 120, loss: 0.0008133051451295614
step: 130, loss: 0.007137095555663109
step: 140, loss: 8.259912283392623e-05
step: 150, loss: 0.00013490211858879775
step: 160, loss: 0.00011161140719195828
step: 170, loss: 0.0061231632716953754
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.8123393316195373, f1=0.798994974874372,
best_f1=0.8126520681265207
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008046521106734872
step: 10, loss: 0.000885374320205301
step: 20, loss: 0.0021679578348994255
step: 30, loss: 0.0015218287007883191
step: 40, loss: 0.0004043675435241312
step: 50, loss: 0.021984094753861427
step: 60, loss: 0.00020285752543713897
step: 70, loss: 0.0053541399538517
step: 80, loss: 0.0032797129824757576
step: 90, loss: 0.000690885994117707
step: 100, loss: 0.0005883712437935174
step: 110, loss: 0.001598789938725531
step: 120, loss: 0.0031382841989398003
step: 130, loss: 0.0020066003780812025
step: 140, loss: 0.00035694209509529173
step: 150, loss: 0.0002427106664981693
step: 160, loss: 0.018936684355139732
step: 170, loss: 0.00011607591295614839
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.8115183246073299, f1=0.8040712468193385,
best_f1=0.8126520681265207
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0015631834976375103
step: 10, loss: 0.005236473865807056
step: 20, loss: 0.03902760148048401
step: 30, loss: 0.016228795051574707
step: 40, loss: 0.000250098790274933
step: 50, loss: 0.0006532616098411381
step: 60, loss: 8.924585563363507e-05
step: 70, loss: 0.0005406148266047239
step: 80, loss: 0.00039533639210276306
step: 90, loss: 0.0019625560380518436
step: 100, loss: 0.0004872881108894944
step: 110, loss: 0.0001113096805056557
step: 120, loss: 0.03749706596136093
step: 130, loss: 0.0028837851714342833
step: 140, loss: 0.0007305059116333723
step: 150, loss: 0.03101460635662079
step: 160, loss: 0.00032183199073188007
step: 170, loss: 0.005700106266885996
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.8140161725067385, f1=0.7958656330749355,
best_f1=0.8126520681265207
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
1916it [00:04, 420.38it/s]
load_f1 = 0.6696230598669624
real_f1 = 0.5988483685220729
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:11, 367.01it/s]

```

1.3 DA DIRTY

1.3.1 DBLP-ACM - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Dirty/DBLP-ACM \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model \
  --da entry_swap \
  --da span_shuffle \
  --summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-ACM.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Dirty/DBLP-ACM \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

Downloading: 100% 442/442 [00:00<00:00, 91.0kB/s]

Downloading: 100% 232k/232k [00:00<00:00, 341kB/s]

Downloading: 100% 268M/268M [00:07<00:00, 34.0MB/s]

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

enabled	: True
opt_level	: 02
cast_model_type	: torch.float16
patch_torch_functions	: False
keep_batchnorm_fp32	: True

```

master_weights      : True
loss_scale          : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled             : True
opt_level           : 02
cast_model_type     : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights      : True
loss_scale          : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8215144276618958
step: 10, loss: 0.4545573592185974
step: 20, loss: 0.556934654712677
step: 30, loss: 0.36844977736473083
step: 40, loss: 0.13983841240406036
step: 50, loss: 0.16636043787002563
step: 60, loss: 0.05547279864549637
step: 70, loss: 0.13986904919147491
step: 80, loss: 0.20614880323410034
step: 90, loss: 0.07294894009828568
step: 100, loss: 0.02669627033174038
step: 110, loss: 0.08010876923799515
step: 120, loss: 0.050120770931243896
step: 130, loss: 0.00474143261089921
step: 140, loss: 0.08038301020860672
step: 150, loss: 0.12880876660346985
step: 160, loss: 0.14956296980381012
step: 170, loss: 0.012383573688566685
step: 180, loss: 0.009633387438952923
step: 190, loss: 0.016075557097792625
step: 200, loss: 0.018266264349222183

```

```

step: 210, loss: 0.019829601049423218
step: 220, loss: 0.011426647193729877
step: 230, loss: 0.018598318099975586
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9599999999999999, f1=0.9686800894854586,
best_f1=0.9686800894854586
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04904277250170708
step: 10, loss: 0.017000436782836914
step: 20, loss: 0.005078254733234644
step: 30, loss: 0.0009468213538639247
step: 40, loss: 0.00381667073816061
step: 50, loss: 0.0026866886764764786
step: 60, loss: 0.0026476357597857714
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 70, loss: 0.05250916630029678
step: 80, loss: 0.004067832138389349
step: 90, loss: 0.030309051275253296
step: 100, loss: 0.10222107917070389
step: 110, loss: 0.10321833938360214
step: 120, loss: 0.004167001694440842
step: 130, loss: 0.21554453670978546
step: 140, loss: 0.23645606637001038
step: 150, loss: 0.00858320016413927
step: 160, loss: 0.028286652639508247
step: 170, loss: 0.004601581022143364
step: 180, loss: 0.06839162111282349
step: 190, loss: 0.141262948513031
step: 200, loss: 0.0034933427814394236
step: 210, loss: 0.22528238594532013
step: 220, loss: 0.003035471774637699
step: 230, loss: 0.03415624052286148
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9688195991091313, f1=0.9678848283499446,

```

```

best_f1=0.9678848283499446
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10950607806444168
step: 10, loss: 0.007354653440415859
step: 20, loss: 0.007924736477434635
step: 30, loss: 0.018934199586510658
step: 40, loss: 0.011898143216967583
step: 50, loss: 0.011465534567832947
step: 60, loss: 0.001970105804502964
step: 70, loss: 0.004077926278114319
step: 80, loss: 0.0006674419273622334
step: 90, loss: 0.0031178726349025965
step: 100, loss: 0.0006910993834026158
step: 110, loss: 0.0016363344620913267
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 120, loss: 0.22097735106945038
step: 130, loss: 0.002024912042543292
step: 140, loss: 0.0051058558747172356
step: 150, loss: 0.00546356663107872
step: 160, loss: 0.07366117089986801
step: 170, loss: 0.01015717163681984
step: 180, loss: 0.0031237967777997255
step: 190, loss: 0.0019110069843009114
step: 200, loss: 0.005341139622032642
step: 210, loss: 0.057941097766160965
step: 220, loss: 0.004420869518071413
step: 230, loss: 0.03375132381916046
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.960272417707151, f1=0.9740112994350283,
best_f1=0.9678848283499446
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014020510949194431
step: 10, loss: 0.0014905462739989161
step: 20, loss: 0.0006021009758114815
step: 30, loss: 0.0010434978175908327
step: 40, loss: 0.002357022138312459
step: 50, loss: 0.002189503749832511
step: 60, loss: 0.0018705996917560697
step: 70, loss: 0.005248654633760452
step: 80, loss: 0.1357276886701584
step: 90, loss: 0.12413673102855682
step: 100, loss: 0.042638592422008514

```

```

step: 110, loss: 0.015411138534545898
step: 120, loss: 0.018022459000349045
step: 130, loss: 0.006703808903694153
step: 140, loss: 0.0009632328292354941
step: 150, loss: 0.0021733904723078012
step: 160, loss: 0.004868456162512302
step: 170, loss: 0.0013867742381989956
step: 180, loss: 0.06091303750872612
step: 190, loss: 0.001292467350140214
step: 200, loss: 0.05245714262127876
step: 210, loss: 0.0035349209792912006
step: 220, loss: 0.0018609932158142328
step: 230, loss: 0.0006756446673534811
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.9680968096809681, f1=0.9636163175303197,
best_f1=0.9678848283499446
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009646786493249238
step: 10, loss: 0.0011844971450045705
step: 20, loss: 0.00055788888130337
step: 30, loss: 0.0009657407645136118
step: 40, loss: 0.0007120961672626436
step: 50, loss: 0.0005152971716597676
step: 60, loss: 0.0006931059178896248
step: 70, loss: 0.000810437195468694
step: 80, loss: 0.0051773469895124435
step: 90, loss: 0.014202692545950413
step: 100, loss: 0.002260650508105755
step: 110, loss: 0.0007620026008225977
step: 120, loss: 0.03909842669963837
step: 130, loss: 0.08579552918672562
step: 140, loss: 0.0011966719757765532
step: 150, loss: 0.0003748386516235769
step: 160, loss: 0.0074340119026601315
step: 170, loss: 0.07318718731403351
step: 180, loss: 0.0011431240709498525
step: 190, loss: 0.0008040482643991709
step: 200, loss: 0.0005343059892766178
step: 210, loss: 0.05931861698627472
step: 220, loss: 0.0011892120819538832
step: 230, loss: 0.0030202013440430164
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```



```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.9705882352941176, f1=0.9705882352941176,
best_f1=0.9705882352941176
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009912493405863643
step: 10, loss: 0.0004230802005622536
step: 20, loss: 0.000765649660024792
step: 30, loss: 0.002104911021888256
step: 40, loss: 0.002547740237787366
step: 50, loss: 0.029858490452170372
step: 60, loss: 0.012594879604876041
step: 70, loss: 0.000351245776982978
step: 80, loss: 0.0013226495357230306
step: 90, loss: 0.00032537459628656507
step: 100, loss: 0.001538632670417428
step: 110, loss: 0.02289450541138649
step: 120, loss: 0.00023761243210174143
step: 130, loss: 0.0023610289208590984
step: 140, loss: 0.00033458968391641974
step: 150, loss: 0.00031014703563414514
step: 160, loss: 0.0014898708323016763
step: 170, loss: 0.00017821193614508957
step: 180, loss: 0.00021437066607177258
step: 190, loss: 0.035398777574300766
step: 200, loss: 0.004097762517631054
step: 210, loss: 0.025702953338623047
step: 220, loss: 0.001517361612059176
step: 230, loss: 0.0007725877221673727
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.967670011148272, f1=0.9740698985343857,
best_f1=0.9705882352941176
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07850994169712067
step: 10, loss: 0.0021170154213905334
step: 20, loss: 0.002915340941399336
step: 30, loss: 0.0009695274056866765
step: 40, loss: 0.0022628167644143105
step: 50, loss: 0.0012326971627771854
step: 60, loss: 0.04958590865135193

```

```

step: 70, loss: 0.0017502546543255448
step: 80, loss: 0.00026661274023354053
step: 90, loss: 0.0015999570023268461
step: 100, loss: 0.002987618325278163
step: 110, loss: 0.009065653197467327
step: 120, loss: 0.0043430207297205925
step: 130, loss: 0.001187117537483573
step: 140, loss: 0.00012212360161356628
step: 150, loss: 0.019025448709726334
step: 160, loss: 0.002662014914676547
step: 170, loss: 0.0003176229365635663
step: 180, loss: 0.00038618993130512536
step: 190, loss: 0.00041523281834088266
step: 200, loss: 0.0012110761599615216
step: 210, loss: 0.00017305328219663352
step: 220, loss: 0.0012599703622981906
step: 230, loss: 0.0366075336933136
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.9720670391061451, f1=0.9688195991091313,
best_f1=0.9688195991091313
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012777434661984444
step: 10, loss: 0.0197280403226614
step: 20, loss: 0.00025991376605816185
step: 30, loss: 0.000915017444640398
step: 40, loss: 0.00519981887191534
step: 50, loss: 0.0029098677914589643
step: 60, loss: 0.0002923669817391783
step: 70, loss: 0.0003974264254793525
step: 80, loss: 0.0017403365345671773
step: 90, loss: 0.0005669400561600924
step: 100, loss: 0.00021612337150145322
step: 110, loss: 0.0010181961115449667
step: 120, loss: 0.00027913699159398675
step: 130, loss: 0.00609827134758234
step: 140, loss: 0.00010881706839427352
step: 150, loss: 0.0002736049355007708
step: 160, loss: 0.004279625602066517
step: 170, loss: 0.0003562620549928397
step: 180, loss: 0.000398533942643553

```

```

step: 190, loss: 0.00032790849218145013
step: 200, loss: 0.001949045923538506
step: 210, loss: 0.0001696801627986133
step: 220, loss: 0.0002088466426357627
step: 230, loss: 0.04289080947637558
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 8: dev_f1=0.9732739420935412, f1=0.9689578713968958,
best_f1=0.9689578713968958
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00021755903435405344
step: 10, loss: 0.0028510678093880415
step: 20, loss: 0.0005574465030804276
step: 30, loss: 0.0002614708209875971
step: 40, loss: 0.012953240424394608
step: 50, loss: 0.0001890950370579958
step: 60, loss: 0.0010144442785531282
step: 70, loss: 0.00034631832386367023
step: 80, loss: 0.024501629173755646
step: 90, loss: 0.005329558625817299
step: 100, loss: 0.0020987465977668762
step: 110, loss: 0.00010668991308193654
step: 120, loss: 0.06124308705329895
step: 130, loss: 0.00020564089936669916
step: 140, loss: 0.019428884610533714
step: 150, loss: 7.430696132360026e-05
step: 160, loss: 0.001615568296983838
step: 170, loss: 0.000126493614516221
step: 180, loss: 0.00034225612762384117
step: 190, loss: 0.0004680771962739527
step: 200, loss: 0.0002191202947869897
step: 210, loss: 0.00010799482697620988
step: 220, loss: 0.030008848756551743
step: 230, loss: 0.00041944146505557
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)

```

```

epoch 9: dev_f1=0.9733333333333333, f1=0.9700332963374029,
best_f1=0.9700332963374029
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00020034828048665076
step: 10, loss: 0.00021693225426133722
step: 20, loss: 0.00012629598495550454
step: 30, loss: 0.00018368834571447223
step: 40, loss: 0.00015909776266198605
step: 50, loss: 0.009916841983795166
step: 60, loss: 0.04715656861662865
step: 70, loss: 0.00039350317092612386
step: 80, loss: 0.00082389375893339912
step: 90, loss: 0.00014456511416938156
step: 100, loss: 0.00014894887863192707
step: 110, loss: 0.00014613213716074824
step: 120, loss: 0.03769103065133095
step: 130, loss: 0.00013665377628058195
step: 140, loss: 0.018543757498264313
step: 150, loss: 0.0008613447425886989
step: 160, loss: 0.00022095930762588978
step: 170, loss: 0.0001322418829659
step: 180, loss: 0.00023063569096848369
step: 190, loss: 0.00021931531955488026
step: 200, loss: 0.0017060647951439023
step: 210, loss: 0.0001280012947972864
step: 220, loss: 0.0002779494971036911
step: 230, loss: 0.0002242999617010355
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9732739420935412, f1=0.9699666295884317,
best_f1=0.9700332963374029
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00010447353270137683
step: 10, loss: 0.00010636196384439245
step: 20, loss: 0.00026567207532934844
step: 30, loss: 0.00043491978431120515
step: 40, loss: 0.00765657564625144
step: 50, loss: 0.00041936017805710435
step: 60, loss: 0.00020994078658986837
step: 70, loss: 0.0001564050035085529
step: 80, loss: 0.0005663397023454309
step: 90, loss: 8.729277033125982e-05
step: 100, loss: 0.00011501980770844966

```

```

step: 110, loss: 0.0005689574172720313
step: 120, loss: 0.001035580295138061
step: 130, loss: 0.0001341021416010335
step: 140, loss: 0.0001260208518942818
step: 150, loss: 8.781749056652188e-05
step: 160, loss: 0.00011933534551644698
step: 170, loss: 0.005751394666731358
step: 180, loss: 0.00025204019038937986
step: 190, loss: 0.0001470356510253623
step: 200, loss: 0.00011939105024794117
step: 210, loss: 5.7958110119216144e-05
step: 220, loss: 9.48069864534773e-05
step: 230, loss: 0.006874362006783485
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9731543624161074, f1=0.9730941704035874,
best_f1=0.9700332963374029
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00020413807942532003
step: 10, loss: 0.00010874738654820248
step: 20, loss: 8.891112520359457e-05
step: 30, loss: 0.0006781970732845366
step: 40, loss: 5.467670052894391e-05
step: 50, loss: 9.014594252221286e-05
step: 60, loss: 0.010924112051725388
step: 70, loss: 6.824792217230424e-05
step: 80, loss: 4.053639349876903e-05
step: 90, loss: 0.00011771926801884547
step: 100, loss: 4.624998473445885e-05
step: 110, loss: 6.279581430135295e-05
step: 120, loss: 7.449925760738552e-05
step: 130, loss: 0.0002075332449749112
step: 140, loss: 8.347658877028152e-05
step: 150, loss: 4.30016552854795e-05
step: 160, loss: 0.00014771358110010624
step: 170, loss: 0.00021555108833126724
step: 180, loss: 0.0001141406100941822
step: 190, loss: 0.0002143378951586783
step: 200, loss: 0.007210922427475452
step: 210, loss: 5.471328040584922e-05
step: 220, loss: 0.0039665414951741695
step: 230, loss: 0.00026764735230244696
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```

```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 12: dev_f1=0.9742441209406495, f1=0.9730337078651685,
best_f1=0.9730337078651685
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.406082633882761e-05
step: 10, loss: 0.0021067389752715826
step: 20, loss: 7.668752368772402e-05
step: 30, loss: 0.0015854832017794251
step: 40, loss: 0.0002037670637946576
step: 50, loss: 0.00022843718761578202
step: 60, loss: 5.090959894005209e-05
step: 70, loss: 5.804547981824726e-05
step: 80, loss: 8.011599857127294e-05
step: 90, loss: 4.5594730181619525e-05
step: 100, loss: 0.00014344227383844554
step: 110, loss: 9.876584226731211e-05
step: 120, loss: 6.043279427103698e-05
step: 130, loss: 0.00011274463031440973
step: 140, loss: 0.001925933756865561
step: 150, loss: 0.002788200508803129
step: 160, loss: 5.61116321478039e-05
step: 170, loss: 7.14725028956309e-05
step: 180, loss: 0.00033430219627916813
step: 190, loss: 0.0003922871546819806
step: 200, loss: 7.529901631642133e-05
step: 210, loss: 0.00014491967158392072
step: 220, loss: 0.0010267229517921805
step: 230, loss: 4.257392720319331e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9741282339707535, f1=0.9708520179372198,
best_f1=0.9730337078651685
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.027575338725e-05
step: 10, loss: 3.729803938767873e-05
step: 20, loss: 0.0008766390383243561
step: 30, loss: 8.865522249834612e-05
step: 40, loss: 3.609997656894848e-05
step: 50, loss: 0.00011783956870203838
step: 60, loss: 3.786582237808034e-05

```

```

step: 70, loss: 7.990236190380529e-05
step: 80, loss: 0.00014724326319992542
step: 90, loss: 0.00020288070663809776
step: 100, loss: 0.006814114283770323
step: 110, loss: 0.0003826561151072383
step: 120, loss: 4.1728559153852984e-05
step: 130, loss: 8.960292325355113e-05
step: 140, loss: 0.00026908560539595783
step: 150, loss: 9.052736277226359e-05
step: 160, loss: 3.079591624555178e-05
step: 170, loss: 0.0006938831065781415
step: 180, loss: 7.85378142609261e-05
step: 190, loss: 0.0002911364135798067
step: 200, loss: 9.166976815322414e-05
step: 210, loss: 0.011931097134947777
step: 220, loss: 0.0014170060167089105
step: 230, loss: 2.294741898367647e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 14: dev_f1=0.9751131221719457, f1=0.972972972972973,
best_f1=0.972972972972973
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00011449351586634293
step: 10, loss: 8.186609920812771e-05
step: 20, loss: 6.2067192629911e-05
step: 30, loss: 9.696112101664767e-05
step: 40, loss: 0.00012546111247502267
step: 50, loss: 9.501275781076401e-05
step: 60, loss: 4.572694524540566e-05
step: 70, loss: 0.00017852822202257812
step: 80, loss: 0.005979168228805065
step: 90, loss: 2.995043723785784e-05
step: 100, loss: 6.030826625647023e-05
step: 110, loss: 6.773544737370685e-05
step: 120, loss: 6.649921124335378e-05
step: 130, loss: 7.048757834127173e-05
step: 140, loss: 0.013194456696510315
step: 150, loss: 0.00010965048568323255
step: 160, loss: 5.947469617240131e-05
step: 170, loss: 3.319121969980188e-05
step: 180, loss: 5.4734548029955477e-05

```

```

step: 190, loss: 0.000285079178865999
step: 200, loss: 3.506093344185501e-05
step: 210, loss: 0.004266345873475075
step: 220, loss: 0.001818401156924665
step: 230, loss: 6.665467662969604e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9742441209406495, f1=0.9697648376259798,
best_f1=0.972972972972973
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2473it [00:10, 227.52it/s]
load_f1 = 0.975609756097561
real_f1 = 0.9733924611973392
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:17, 247.59it/s]

```


1.3.2 DBLP-GoogleScholar - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Dirty/DBLP-GoogleScholar \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model \
  --da span_shuffle \
  --da attr_del \
  --summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-GoogleScholar/.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Dirty/DBLP-GoogleScholar \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True
```

```

opt_level                : 02
cast_model_type          : torch.float16
patch_torch_functions    : False
keep_batchnorm_fp32      : True
master_weights           : True
loss_scale                : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.7966650128364563
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4535463750362396
step: 20, loss: 0.48909905552864075
step: 30, loss: 0.42566248774528503
step: 40, loss: 0.3015522062778473
step: 50, loss: 0.2031581848859787
step: 60, loss: 0.16260962188243866
step: 70, loss: 0.10154974460601807
step: 80, loss: 0.1864793598651886
step: 90, loss: 0.09593465179204941
step: 100, loss: 0.2115669846534729
step: 110, loss: 0.10305113345384598
step: 120, loss: 0.024174364283680916
step: 130, loss: 0.04026225954294205
step: 140, loss: 0.23750001192092896
step: 150, loss: 0.027972521260380745
step: 160, loss: 0.23839549720287323
step: 170, loss: 0.13333432376384735
step: 180, loss: 0.1496521681547165
step: 190, loss: 0.05943674594163895
step: 200, loss: 0.1088145300745964
step: 210, loss: 0.11969156563282013
step: 220, loss: 0.10860315710306168
step: 230, loss: 0.07408715784549713
step: 240, loss: 0.1311013102531433

```

```

step: 250, loss: 0.04996468499302864
step: 260, loss: 0.03700567036867142
step: 270, loss: 0.026625823229551315
step: 280, loss: 0.10580437630414963
step: 290, loss: 0.06096607819199562
step: 300, loss: 0.04769781231880188
step: 310, loss: 0.0181167833507061
step: 320, loss: 0.05839037895202637
step: 330, loss: 0.21214821934700012
step: 340, loss: 0.23350732028484344
step: 350, loss: 0.09796424210071564
step: 360, loss: 0.0719894990324974
step: 370, loss: 0.12790396809577942
step: 380, loss: 0.35628965497016907
step: 390, loss: 0.06100636348128319
step: 400, loss: 0.024739522486925125
step: 410, loss: 0.06756880134344101
step: 420, loss: 0.008971279487013817
step: 430, loss: 0.12383782118558884
step: 440, loss: 0.10313039273023605
step: 450, loss: 0.056895408779382706
step: 460, loss: 0.021719468757510185
step: 470, loss: 0.07777491956949234
step: 480, loss: 0.2914702296257019
step: 490, loss: 0.021463647484779358
step: 500, loss: 0.01044117659330368
step: 510, loss: 0.06882673501968384
step: 520, loss: 0.1287781149148941
step: 530, loss: 0.06716044247150421
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9310183012670109, f1=0.923221855864343,
best_f1=0.923221855864343
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11929343640804291
step: 10, loss: 0.14157885313034058
step: 20, loss: 0.21611285209655762
step: 30, loss: 0.03034740686416626
step: 40, loss: 0.006721616722643375
step: 50, loss: 0.09623957425355911
step: 60, loss: 0.1192820742726326

```

step: 70, loss: 0.1336098611354828
step: 80, loss: 0.03712771087884903
step: 90, loss: 0.013458305038511753
step: 100, loss: 0.46615082025527954
step: 110, loss: 0.01766500249505043
step: 120, loss: 0.17544345557689667
step: 130, loss: 0.0341181606054306
step: 140, loss: 0.09015274792909622
step: 150, loss: 0.07724901288747787
step: 160, loss: 0.04106111824512482
step: 170, loss: 0.1284203678369522
step: 180, loss: 0.008619451895356178
step: 190, loss: 0.01490139402449131
step: 200, loss: 0.028723087161779404
step: 210, loss: 0.014791366644203663
step: 220, loss: 0.1883404552936554
step: 230, loss: 0.006707922089844942
step: 240, loss: 0.2527599632740021
step: 250, loss: 0.01517900824546814
step: 260, loss: 0.04045334830880165
step: 270, loss: 0.08794307708740234
step: 280, loss: 0.10198426246643066
step: 290, loss: 0.07766152173280716
step: 300, loss: 0.0875604897737503
step: 310, loss: 0.05690527707338333
step: 320, loss: 0.04533272609114647
step: 330, loss: 0.04796525463461876
step: 340, loss: 0.0900651216506958
step: 350, loss: 0.07247161865234375
step: 360, loss: 0.03343803808093071
step: 370, loss: 0.009904278442263603
step: 380, loss: 0.07945746928453445
step: 390, loss: 0.06698519736528397
step: 400, loss: 0.08283839374780655
step: 410, loss: 0.0015164330834522843
step: 420, loss: 0.01943710632622242
step: 430, loss: 0.017780421301722527
step: 440, loss: 0.025424089282751083
step: 450, loss: 0.011768425814807415
step: 460, loss: 0.24607083201408386
step: 470, loss: 0.02086673118174076
step: 480, loss: 0.20300517976284027
step: 490, loss: 0.03831809014081955
step: 500, loss: 0.03998025506734848
step: 510, loss: 0.0899917408823967
step: 520, loss: 0.12743346393108368
step: 530, loss: 0.08915779739618301
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9352319706017456, f1=0.9329685362517101,
best_f1=0.9329685362517101
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.016725558787584305
step: 10, loss: 0.09897191822528839
step: 20, loss: 0.23295331001281738
step: 30, loss: 0.27650943398475647
step: 40, loss: 0.017687641084194183
step: 50, loss: 0.040372904390096664
step: 60, loss: 0.0048089781776070595
step: 70, loss: 0.022912081331014633
step: 80, loss: 0.0005854148766957223
step: 90, loss: 0.19365017116069794
step: 100, loss: 0.022178716957569122
step: 110, loss: 0.012595443986356258
step: 120, loss: 0.005963566713035107
step: 130, loss: 0.021753421053290367
step: 140, loss: 0.060481827706098557
step: 150, loss: 0.0179400946944952
step: 160, loss: 0.006778695620596409
step: 170, loss: 0.020499132573604584
step: 180, loss: 0.005262086633592844
step: 190, loss: 0.004531098995357752
step: 200, loss: 0.0062440247274935246
step: 210, loss: 0.03376224637031555
step: 220, loss: 0.03243016451597214
step: 230, loss: 0.06880607455968857
step: 240, loss: 0.04656228423118591
step: 250, loss: 0.006929093040525913
step: 260, loss: 0.038161300122737885
step: 270, loss: 0.005793389864265919
step: 280, loss: 0.016931898891925812
step: 290, loss: 0.07480453699827194
step: 300, loss: 0.12330451607704163
step: 310, loss: 0.17261695861816406
step: 320, loss: 0.10792764276266098
step: 330, loss: 0.0035242754966020584
step: 340, loss: 0.002673092298209667
step: 350, loss: 0.04862048104405403
step: 360, loss: 0.007969927042722702

```

```
step: 370, loss: 0.0020313060376793146
step: 380, loss: 0.009751327335834503
step: 390, loss: 0.022692766040563583
step: 400, loss: 0.05790356174111366
step: 410, loss: 0.012551468797028065
step: 420, loss: 0.03126018866896629
step: 430, loss: 0.046565569937229156
step: 440, loss: 0.033787790685892105
step: 450, loss: 0.07083127647638321
step: 460, loss: 0.11934570223093033
step: 470, loss: 0.007427125237882137
step: 480, loss: 0.004474115092307329
step: 490, loss: 0.05358167737722397
step: 500, loss: 0.047258276492357254
step: 510, loss: 0.0051672980189323425
step: 520, loss: 0.0039966111071407795
step: 530, loss: 0.0649259015917778
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.9286043298019346, f1=0.9296296296296297,
best_f1=0.9329685362517101
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01080924179404974
step: 10, loss: 0.005525710992515087
step: 20, loss: 0.0475163534283638
step: 30, loss: 0.002142974641174078
step: 40, loss: 0.004348484333604574
step: 50, loss: 0.19644810259342194
step: 60, loss: 0.0039205425418913364
step: 70, loss: 0.0029993776697665453
step: 80, loss: 0.005011203233152628
step: 90, loss: 0.04893823340535164
step: 100, loss: 0.009316311217844486
step: 110, loss: 0.003014638088643551
step: 120, loss: 0.010744879953563213
step: 130, loss: 0.10076679289340973
step: 140, loss: 0.016369983553886414
step: 150, loss: 0.014324125833809376
step: 160, loss: 0.0017097527161240578
step: 170, loss: 0.009853915311396122
step: 180, loss: 0.02150495909154415
step: 190, loss: 0.0011446747230365872
step: 200, loss: 0.0004387479566503316
step: 210, loss: 0.01631927117705345
step: 220, loss: 0.02843417041003704
```

```
step: 230, loss: 0.15431363880634308
step: 240, loss: 0.010431472212076187
step: 250, loss: 0.004181009251624346
step: 260, loss: 0.15261322259902954
step: 270, loss: 0.013086779974400997
step: 280, loss: 0.011566292494535446
step: 290, loss: 0.19869989156723022
step: 300, loss: 0.0017750711413100362
step: 310, loss: 0.0032446407712996006
step: 320, loss: 0.022136516869068146
step: 330, loss: 0.008646017871797085
step: 340, loss: 0.05801645666360855
step: 350, loss: 0.03916071355342865
step: 360, loss: 0.005517083220183849
step: 370, loss: 0.06316664814949036
step: 380, loss: 0.03985381871461868
step: 390, loss: 0.05307302623987198
step: 400, loss: 0.016291795298457146
step: 410, loss: 0.02671700157225132
step: 420, loss: 0.00919979065656662
step: 430, loss: 0.009862513281404972
step: 440, loss: 0.09325719624757767
step: 450, loss: 0.025615990161895752
step: 460, loss: 0.004242536146193743
step: 470, loss: 0.0061780535615980625
step: 480, loss: 0.022337883710861206
step: 490, loss: 0.012986625544726849
step: 500, loss: 0.05913834273815155
step: 510, loss: 0.03650300204753876
step: 520, loss: 0.026987535879015923
step: 530, loss: 0.00927529763430357
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.9319419237749547, f1=0.9289568345323741,
best_f1=0.9329685362517101
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003965152893215418
step: 10, loss: 0.0311654731631279
step: 20, loss: 0.0037418671417981386
step: 30, loss: 0.0075628794729709625
step: 40, loss: 0.14771205186843872
step: 50, loss: 0.03581734746694565
step: 60, loss: 0.0035603828728199005
step: 70, loss: 0.0007807879592292011
step: 80, loss: 0.0007679041591472924
```

step: 90, loss: 0.10266561061143875
step: 100, loss: 0.0028311805799603462
step: 110, loss: 0.016395531594753265
step: 120, loss: 0.04811226576566696
step: 130, loss: 0.0013453150168061256
step: 140, loss: 0.010018120519816875
step: 150, loss: 0.010033300146460533
step: 160, loss: 0.04867326468229294
step: 170, loss: 0.04675145819783211
step: 180, loss: 0.045990485697984695
step: 190, loss: 0.009610706940293312
step: 200, loss: 0.05430722236633301
step: 210, loss: 0.0054698544554412365
step: 220, loss: 0.0009417671244591475
step: 230, loss: 0.0005563572631217539
step: 240, loss: 0.015288930386304855
step: 250, loss: 0.0009673485765233636
step: 260, loss: 0.009490138851106167
step: 270, loss: 0.05426454916596413
step: 280, loss: 0.0423271581530571
step: 290, loss: 0.010470705106854439
step: 300, loss: 0.025578023865818977
step: 310, loss: 0.0012097549624741077
step: 320, loss: 0.06607324630022049
step: 330, loss: 0.009700744412839413
step: 340, loss: 0.010347777046263218
step: 350, loss: 0.022095169872045517
step: 360, loss: 0.06600669771432877
step: 370, loss: 0.0009615019080229104
step: 380, loss: 0.04940079152584076
step: 390, loss: 0.005176727660000324
step: 400, loss: 0.10324946790933609
step: 410, loss: 0.004075631033629179
step: 420, loss: 0.022314896807074547
step: 430, loss: 0.008691041730344296
step: 440, loss: 0.01328309066593647
step: 450, loss: 0.06276369094848633
step: 460, loss: 0.023841043934226036
step: 470, loss: 0.0024705715477466583
step: 480, loss: 0.001013101194985211
step: 490, loss: 0.048774611204862595
step: 500, loss: 0.011368023231625557
step: 510, loss: 0.0009797548409551382
step: 520, loss: 0.004258425906300545
step: 530, loss: 0.04708544537425041

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

warnings.warn("An input tensor was not cuda.")


```
epoch 5: dev_f1=0.9322892676186089, f1=0.932475884244373,  
best_f1=0.9329685362517101  
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:  
An input tensor was not cuda.  
    warnings.warn("An input tensor was not cuda.")  
step: 0, loss: 0.0045129586942493916  
step: 10, loss: 0.0014210959197953343  
step: 20, loss: 0.0002729099942371249  
step: 30, loss: 0.004041708540171385  
step: 40, loss: 0.0028970607090741396  
step: 50, loss: 0.007338109891861677  
step: 60, loss: 0.0033552648965269327  
step: 70, loss: 0.03220387175679207  
step: 80, loss: 0.0064064254984259605  
step: 90, loss: 0.08928985893726349  
step: 100, loss: 0.030597658827900887  
step: 110, loss: 0.007134160026907921  
step: 120, loss: 0.10874960571527481  
step: 130, loss: 0.02338174544274807  
step: 140, loss: 0.006919438019394875  
step: 150, loss: 0.004349390044808388  
step: 160, loss: 0.0004537561326287687  
step: 170, loss: 0.00025792833184823394  
step: 180, loss: 0.003903178498148918  
step: 190, loss: 0.07611721009016037  
step: 200, loss: 0.0002156483824364841  
step: 210, loss: 0.005904064048081636  
step: 220, loss: 0.028837362304329872  
step: 230, loss: 0.00038488334394060075  
step: 240, loss: 0.002496624831110239  
step: 250, loss: 0.0006827439065091312  
step: 260, loss: 0.001367758377455175  
step: 270, loss: 0.0022217403165996075  
step: 280, loss: 0.0018993997946381569  
step: 290, loss: 0.014572251588106155  
step: 300, loss: 0.009230714291334152  
step: 310, loss: 0.020839784294366837  
step: 320, loss: 0.009330296888947487  
step: 330, loss: 0.0683077871799469  
step: 340, loss: 0.045153915882110596  
step: 350, loss: 0.03544801473617554  
step: 360, loss: 0.0003307618317194283  
step: 370, loss: 0.03648269549012184  
step: 380, loss: 0.006605245638638735  
step: 390, loss: 0.1245780736207962  
step: 400, loss: 0.012898696586489677  
step: 410, loss: 0.00013446714729070663  
step: 420, loss: 0.04963820427656174
```

```
step: 430, loss: 0.001072774059139192
step: 440, loss: 0.001597419730387628
step: 450, loss: 0.0034454395063221455
step: 460, loss: 0.0018977702129632235
step: 470, loss: 0.030553841963410378
step: 480, loss: 0.014054610393941402
step: 490, loss: 0.000681289762724191
step: 500, loss: 0.0036306879483163357
step: 510, loss: 0.0001248977641807869
step: 520, loss: 0.0035670408979058266
step: 530, loss: 0.008064083755016327
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.927315914489311, f1=0.9219047619047618,
best_f1=0.9329685362517101
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003588959341868758
step: 10, loss: 0.012764987535774708
step: 20, loss: 0.0028701021801680326
step: 30, loss: 0.007045058533549309
step: 40, loss: 0.0010721477447077632
step: 50, loss: 0.0015399240655824542
step: 60, loss: 0.03012661077082157
step: 70, loss: 0.001780361053533852
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 80, loss: 0.0003543384082149714
step: 90, loss: 0.0007765181944705546
step: 100, loss: 0.008863239549100399
step: 110, loss: 0.00042548932833597064
step: 120, loss: 0.0009494024561718106
step: 130, loss: 0.003449120558798313
step: 140, loss: 0.014720575883984566
step: 150, loss: 0.0002725528320297599
step: 160, loss: 0.0013243148569017649
step: 170, loss: 0.002873094752430916
step: 180, loss: 0.0029427011031657457
step: 190, loss: 0.0001501437945989892
step: 200, loss: 0.013008184731006622
step: 210, loss: 0.0011989653576165438
step: 220, loss: 0.0003111807454843074
step: 230, loss: 0.004053354728966951
step: 240, loss: 0.015340717509388924
step: 250, loss: 0.004363623913377523
step: 260, loss: 0.012164395302534103
step: 270, loss: 0.0004721196601167321
```

```

step: 280, loss: 0.06569754332304001
step: 290, loss: 0.0012012694496661425
step: 300, loss: 0.0008530043996870518
step: 310, loss: 0.006390595342963934
step: 320, loss: 0.08170083910226822
step: 330, loss: 0.0004815045977011323
step: 340, loss: 0.02867802418768406
step: 350, loss: 0.009881867095828056
step: 360, loss: 0.00432220846414566
step: 370, loss: 0.024174442514777184
step: 380, loss: 0.0024367799051105976
step: 390, loss: 0.0019322083098813891
step: 400, loss: 0.0006527036894112825
step: 410, loss: 0.0019388914806768298
step: 420, loss: 0.0015142239863052964
step: 430, loss: 0.01424350030720234
step: 440, loss: 0.06217473745346069
step: 450, loss: 0.004262377507984638
step: 460, loss: 0.0022129800636321306
step: 470, loss: 0.04084330052137375
step: 480, loss: 0.0004073175950907171
step: 490, loss: 0.0031183187384158373
step: 500, loss: 0.0003141143242828548
step: 510, loss: 0.0014683038461953402
step: 520, loss: 0.005271553993225098
step: 530, loss: 0.0002950764319393784
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.936768149882904, f1=0.9286047596826879,
best_f1=0.9286047596826879
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.020040743052959442
step: 10, loss: 0.0107731306925416
step: 20, loss: 0.046374667435884476
step: 30, loss: 0.016805199906229973
step: 40, loss: 0.008582794107496738
step: 50, loss: 0.0029052807949483395
step: 60, loss: 0.009704113937914371
step: 70, loss: 0.0009193435544148088
step: 80, loss: 0.00018952299433294684
step: 90, loss: 3.525861757225357e-05

```

```
step: 100, loss: 0.02827065996825695
step: 110, loss: 0.0013598694931715727
step: 120, loss: 0.000777740846388042
step: 130, loss: 0.022103218361735344
step: 140, loss: 4.9913989641936496e-05
step: 150, loss: 0.0007116114138625562
step: 160, loss: 0.0007117212517186999
step: 170, loss: 0.015283452346920967
step: 180, loss: 0.00021564045164268464
step: 190, loss: 0.0038708020001649857
step: 200, loss: 0.001603389740921557
step: 210, loss: 0.00023724217317067087
step: 220, loss: 0.0013687886530533433
step: 230, loss: 0.0013161861570551991
step: 240, loss: 0.00015059187717270106
step: 250, loss: 0.0017249706434085965
step: 260, loss: 0.22175626456737518
step: 270, loss: 0.0006400386337190866
step: 280, loss: 0.006683481391519308
step: 290, loss: 0.07615388184785843
step: 300, loss: 0.0004105776606593281
step: 310, loss: 0.05496963858604431
step: 320, loss: 0.0008459473610855639
step: 330, loss: 0.004660036880522966
step: 340, loss: 0.0009183523943647742
step: 350, loss: 0.026860486716032028
step: 360, loss: 0.003413701895624399
step: 370, loss: 0.0010028679389506578
step: 380, loss: 0.025397509336471558
step: 390, loss: 0.018871624022722244
step: 400, loss: 0.07630445063114166
step: 410, loss: 0.0038714802358299494
step: 420, loss: 0.00013070082059130073
step: 430, loss: 0.007376582827419043
step: 440, loss: 0.001437750761397183
step: 450, loss: 0.0011244070483371615
step: 460, loss: 0.001095547224394977
step: 470, loss: 0.007786758244037628
step: 480, loss: 0.0002719999465625733
step: 490, loss: 0.0022726517636328936
step: 500, loss: 0.001760933198966086
step: 510, loss: 0.04352978616952896
step: 520, loss: 0.00046817868133075535
step: 530, loss: 0.00029393937438726425
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9360709286047598, f1=0.9327770050996754,
```

```
best_f1=0.9286047596826879
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005468000308610499
step: 10, loss: 0.00017286381626036018
step: 20, loss: 0.0001133179321186617
step: 30, loss: 0.01089127641171217
step: 40, loss: 0.009146896190941334
step: 50, loss: 0.0010980406077578664
step: 60, loss: 0.002891545882448554
step: 70, loss: 0.11088012158870697
step: 80, loss: 0.0017322005005553365
step: 90, loss: 0.03703118488192558
step: 100, loss: 0.002463076962158084
step: 110, loss: 0.016013460233807564
step: 120, loss: 0.000688828295096755
step: 130, loss: 0.0008814035099931061
step: 140, loss: 0.028272047638893127
step: 150, loss: 0.000486370554426685
step: 160, loss: 0.0021095110569149256
step: 170, loss: 0.00011948506289627403
step: 180, loss: 0.015885496512055397
step: 190, loss: 0.0017018560320138931
step: 200, loss: 0.002229435835033655
step: 210, loss: 0.0005412183236330748
step: 220, loss: 0.001024874160066247
step: 230, loss: 0.0017034977208822966
step: 240, loss: 0.00021061298321001232
step: 250, loss: 0.0232545118778944
step: 260, loss: 0.003713438054546714
step: 270, loss: 0.004011192824691534
step: 280, loss: 4.6398734411923215e-05
step: 290, loss: 0.00024957352434284985
step: 300, loss: 0.008701665326952934
step: 310, loss: 5.256418080534786e-05
step: 320, loss: 0.0003544155624695122
step: 330, loss: 0.0002467388694640249
step: 340, loss: 0.000588502618484199
step: 350, loss: 0.00042881115223281085
step: 360, loss: 0.033368587493896484
step: 370, loss: 0.00021656839817296714
step: 380, loss: 0.00014190172078087926
step: 390, loss: 0.00010002950148191303
step: 400, loss: 0.002176683396100998
step: 410, loss: 0.006502910051494837
step: 420, loss: 0.0007193266064859927
step: 430, loss: 3.380123962415382e-05
```

```

step: 440, loss: 9.385306475451216e-05
step: 450, loss: 7.368931255768985e-05
step: 460, loss: 0.00021437402756419033
step: 470, loss: 0.0003684514085762203
step: 480, loss: 0.004607993643730879
step: 490, loss: 0.0012310313759371638
step: 500, loss: 0.018096622079610825
step: 510, loss: 0.022438431158661842
step: 520, loss: 0.0011066162260249257
step: 530, loss: 0.0001485209504608065
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9335863377609108, f1=0.9329556185080263,
best_f1=0.9286047596826879
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005643990356475115
step: 10, loss: 0.000198075344087556
step: 20, loss: 0.0005103567964397371
step: 30, loss: 0.00021974346600472927
step: 40, loss: 0.00027029289049096406
step: 50, loss: 0.001062809256836772
step: 60, loss: 0.001195001183077693
step: 70, loss: 0.009352599270641804
step: 80, loss: 0.00045360528747551143
step: 90, loss: 0.04676869139075279
step: 100, loss: 0.0002557729894760996
step: 110, loss: 0.0008003701223060489
step: 120, loss: 0.00012281957606319338
step: 130, loss: 0.0012828321196138859
step: 140, loss: 0.003803181927651167
step: 150, loss: 9.884274186333641e-05
step: 160, loss: 0.020697321742773056
step: 170, loss: 0.00016752746887505054
step: 180, loss: 0.09466347098350525
step: 190, loss: 0.00098994723521173
step: 200, loss: 0.0003191253053955734
step: 210, loss: 0.00010686515452107415
step: 220, loss: 0.0032440233044326305
step: 230, loss: 6.326880975393578e-05
step: 240, loss: 5.481334665091708e-05
step: 250, loss: 0.0007307829800993204
step: 260, loss: 0.0003877982671838254
step: 270, loss: 0.002921554259955883
step: 280, loss: 9.44958082982339e-05
step: 290, loss: 0.00019213282212149352

```

```

step: 300, loss: 0.05670782923698425
step: 310, loss: 0.001745543908327818
step: 320, loss: 0.0005107152392156422
step: 330, loss: 0.0215174350887537
step: 340, loss: 0.0009578435565344989
step: 350, loss: 0.25060486793518066
step: 360, loss: 0.00048398159560747445
step: 370, loss: 0.005162158515304327
step: 380, loss: 0.00034331646747887135
step: 390, loss: 5.306949242367409e-05
step: 400, loss: 0.002648530760779977
step: 410, loss: 0.00020962739654351026
step: 420, loss: 0.0004944569664075971
step: 430, loss: 0.0002183752367272973
step: 440, loss: 7.234416989376768e-05
step: 450, loss: 0.006235217209905386
step: 460, loss: 0.010799948126077652
step: 470, loss: 0.0002540778659749776
step: 480, loss: 0.0007423268398270011
step: 490, loss: 0.06794021278619766
step: 500, loss: 0.03839423879981041
step: 510, loss: 0.0007917466573417187
step: 520, loss: 0.00044206116581335664
step: 530, loss: 0.00032492284663021564
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9352451433857539, f1=0.9348623853211009,
best_f1=0.9286047596826879
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005282817408442497
step: 10, loss: 0.0013341567246243358
step: 20, loss: 5.958582551102154e-05
step: 30, loss: 0.0008824532851576805
step: 40, loss: 0.0011621251469478011
step: 50, loss: 0.016231972724199295
step: 60, loss: 8.16938845673576e-05
step: 70, loss: 0.00013901224883738905
step: 80, loss: 0.00022913364227861166
step: 90, loss: 0.0017573048826307058
step: 100, loss: 4.443497527972795e-05
step: 110, loss: 0.00020598740957211703
step: 120, loss: 0.0073355622589588165
step: 130, loss: 0.0011603346792981029
step: 140, loss: 0.00016837970179039985
step: 150, loss: 0.00011070910113630816

```

```

step: 160, loss: 0.015903130173683167
step: 170, loss: 0.0006111697293817997
step: 180, loss: 4.501261719269678e-05
step: 190, loss: 0.001653280109167099
step: 200, loss: 0.0025540916249156
step: 210, loss: 2.079419937217608e-05
step: 220, loss: 0.0003146411618217826
step: 230, loss: 0.0007895873277448118
step: 240, loss: 9.602764475857839e-05
step: 250, loss: 0.0005257094744592905
step: 260, loss: 2.2317326511256397e-05
step: 270, loss: 0.002630088943988085
step: 280, loss: 0.00211996934376657
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 290, loss: 0.01697397790849209
step: 300, loss: 0.0038632398936897516
step: 310, loss: 0.08548834919929504
step: 320, loss: 0.02716165967285633
step: 330, loss: 0.0038357670418918133
step: 340, loss: 0.00037989296833984554
step: 350, loss: 0.0013418104499578476
step: 360, loss: 0.0003567308885976672
step: 370, loss: 0.000124235637485981
step: 380, loss: 0.00010566135460976511
step: 390, loss: 0.016957098618149757
step: 400, loss: 0.00012570066610351205
step: 410, loss: 0.004366558976471424
step: 420, loss: 0.008982679806649685
step: 430, loss: 0.009491929784417152
step: 440, loss: 0.0012742204125970602
step: 450, loss: 0.0011484511196613312
step: 460, loss: 0.008594541810452938
step: 470, loss: 0.013396764174103737
step: 480, loss: 0.011299013160169125
step: 490, loss: 0.04024020582437515
step: 500, loss: 0.007588366977870464
step: 510, loss: 3.909496808773838e-05
step: 520, loss: 2.7704018066287972e-05
step: 530, loss: 0.00043226589332334697
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9330210772833724, f1=0.9326473339569691,
best_f1=0.9286047596826879
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000348907255101949

```


step: 10, loss: 5.6807191867846996e-05
step: 20, loss: 3.687256685225293e-05
step: 30, loss: 0.002187195001170039
step: 40, loss: 8.377541234949604e-05
step: 50, loss: 0.00038687512278556824
step: 60, loss: 0.007651906926184893
step: 70, loss: 0.00021108669170644134
step: 80, loss: 0.0014552739448845387
step: 90, loss: 0.02045055851340294
step: 100, loss: 3.250146983191371e-05
step: 110, loss: 0.001658676890656352
step: 120, loss: 4.936443656333722e-05
step: 130, loss: 0.00021850652410648763
step: 140, loss: 0.0008819776121526957
step: 150, loss: 0.0018410307820886374
step: 160, loss: 0.004512842278927565
step: 170, loss: 0.0017919997917488217
step: 180, loss: 0.0023503501433879137
step: 190, loss: 2.2991962396190502e-05
step: 200, loss: 4.2538405978120863e-05
step: 210, loss: 0.0005031364853493869
step: 220, loss: 1.7717282389639877e-05
step: 230, loss: 0.007899188436567783
step: 240, loss: 0.00020380428759381175
step: 250, loss: 0.0001516756892669946
step: 260, loss: 0.0006337399245239794
step: 270, loss: 5.833048635395244e-05
step: 280, loss: 3.4634926123544574e-05
step: 290, loss: 0.009171721525490284
step: 300, loss: 0.0016523739323019981
step: 310, loss: 0.0003930766542907804
step: 320, loss: 0.0011857112403959036
step: 330, loss: 7.62845593271777e-05
step: 340, loss: 0.0002347211557207629
step: 350, loss: 0.006753683555871248
step: 360, loss: 0.0019607169087976217
step: 370, loss: 4.7158191591734067e-05
step: 380, loss: 0.00022645809804089367
step: 390, loss: 8.056065416894853e-05
step: 400, loss: 0.0002824126568157226
step: 410, loss: 0.018367430195212364
step: 420, loss: 0.0004642331041395664
step: 430, loss: 0.0007573309703730047
step: 440, loss: 0.00012856732064392418
step: 450, loss: 7.552612078143284e-05
step: 460, loss: 0.0019561380613595247
step: 470, loss: 3.5759097954723984e-05
step: 480, loss: 8.153674571076408e-05

```

step: 490, loss: 0.0068718320690095425
step: 500, loss: 0.05037939175963402
step: 510, loss: 7.210083276731893e-05
step: 520, loss: 0.005257546901702881
step: 530, loss: 0.0013183716218918562
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9333954354913834, f1=0.9358736059479555,
best_f1=0.9286047596826879
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006366313900798559
step: 10, loss: 0.003896046429872513
step: 20, loss: 0.00485114473849535
step: 30, loss: 0.004755499307066202
step: 40, loss: 3.1537005270365626e-05
step: 50, loss: 0.0001254690287169069
step: 60, loss: 0.0005935936351306736
step: 70, loss: 0.019800862297415733
step: 80, loss: 0.0021093925461173058
step: 90, loss: 0.0018095988780260086
step: 100, loss: 0.00013361421588342637
step: 110, loss: 0.00010782250319607556
step: 120, loss: 3.651483712019399e-05
step: 130, loss: 0.0009789433097466826
step: 140, loss: 4.516123590292409e-05
step: 150, loss: 0.011222383938729763
step: 160, loss: 3.3517371775815263e-05
step: 170, loss: 3.698796354001388e-05
step: 180, loss: 3.4635460906429216e-05
step: 190, loss: 0.0007218465907499194
step: 200, loss: 0.0027345039416104555
step: 210, loss: 0.0023287679068744183
step: 220, loss: 0.0037330465856939554
step: 230, loss: 0.00020548270549625158
step: 240, loss: 2.1792347979499027e-05
step: 250, loss: 0.09946749359369278
step: 260, loss: 4.805468415725045e-05
step: 270, loss: 0.000214413259527646
step: 280, loss: 2.147585473721847e-05
step: 290, loss: 5.3734853281639516e-05
step: 300, loss: 0.00023654584947507828
step: 310, loss: 0.0015577521407976747
step: 320, loss: 0.002787525998428464
step: 330, loss: 0.00035943399416282773
step: 340, loss: 0.005237284582108259

```

```

step: 350, loss: 0.025855377316474915
step: 360, loss: 4.0060669562080875e-05
step: 370, loss: 0.0002199050795752555
step: 380, loss: 0.003768604714423418
step: 390, loss: 3.314294372103177e-05
step: 400, loss: 4.004691800219007e-05
step: 410, loss: 0.0001342051982646808
step: 420, loss: 0.00014458551595453173
step: 430, loss: 9.329464955953881e-05
step: 440, loss: 0.0003599520423449576
step: 450, loss: 0.0035818913020193577
step: 460, loss: 8.20208151708357e-05
step: 470, loss: 0.015192653983831406
step: 480, loss: 0.015319466590881348
step: 490, loss: 0.00012629234697669744
step: 500, loss: 0.00011882128455908969
step: 510, loss: 0.0033922160509973764
step: 520, loss: 4.1550891182851046e-05
step: 530, loss: 0.00017521886911708862
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9363086936308694, f1=0.9381918819188192,
best_f1=0.9286047596826879
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001601717376615852
step: 10, loss: 3.3599553717067465e-05
step: 20, loss: 1.7322294297628105e-05
step: 30, loss: 0.0005578517448157072
step: 40, loss: 0.011853883042931557
step: 50, loss: 0.0002970286295749247
step: 60, loss: 2.2424525013775565e-05
step: 70, loss: 0.006302379071712494
step: 80, loss: 0.0001401634799549356
step: 90, loss: 5.962785144220106e-05
step: 100, loss: 0.0002267521631438285
step: 110, loss: 2.3244776457431726e-05
step: 120, loss: 0.004278737585991621
step: 130, loss: 0.0011848650174215436
step: 140, loss: 0.03489072620868683
step: 150, loss: 0.0006243396201170981
step: 160, loss: 0.001683817245066166
step: 170, loss: 0.0006569166434928775
step: 180, loss: 0.0011439071968197823
step: 190, loss: 0.0019425565842539072
step: 200, loss: 0.00010570493759587407

```

```

step: 210, loss: 0.02537199854850769
step: 220, loss: 0.002070321701467037
step: 230, loss: 0.0018353818450123072
step: 240, loss: 0.001217784360051155
step: 250, loss: 0.001233372138813138
step: 260, loss: 0.00020881003001704812
step: 270, loss: 0.002821533940732479
step: 280, loss: 5.438005246105604e-05
step: 290, loss: 0.025597119703888893
step: 300, loss: 0.0005105934687890112
step: 310, loss: 4.68958642159123e-05
step: 320, loss: 3.6577599530573934e-05
step: 330, loss: 0.00037754126242361963
step: 340, loss: 4.4291289668763056e-05
step: 350, loss: 0.001274734502658248
step: 360, loss: 0.0013799393782392144
step: 370, loss: 0.016155123710632324
step: 380, loss: 2.241016409243457e-05
step: 390, loss: 6.523662159452215e-05
step: 400, loss: 1.4554501831298694e-05
step: 410, loss: 4.57768292108085e-05
step: 420, loss: 5.939474067417905e-05
step: 430, loss: 3.08726703224238e-05
step: 440, loss: 1.8588241800898686e-05
step: 450, loss: 0.0002080948033835739
step: 460, loss: 0.0037579473573714495
step: 470, loss: 0.00017756177112460136
step: 480, loss: 1.6465528460685164e-05
step: 490, loss: 6.19785932940431e-05
step: 500, loss: 0.007468126714229584
step: 510, loss: 6.745409336872399e-05
step: 520, loss: 6.465901242336258e-05
step: 530, loss: 0.0003928240039385855
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9360518999073215, f1=0.939435968562182,
best_f1=0.9286047596826879
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.510437666904181e-05
step: 10, loss: 3.172489596181549e-05
step: 20, loss: 4.849974357057363e-05
step: 30, loss: 0.00034712444175966084
step: 40, loss: 0.033510927110910416
step: 50, loss: 0.00029766367515549064
step: 60, loss: 0.001832168665714562

```

step: 70, loss: 0.00012052110832883045
step: 80, loss: 5.5912336392793804e-05
step: 90, loss: 0.006796527653932571
step: 100, loss: 0.0020458055660128593
step: 110, loss: 0.01277808379381895
step: 120, loss: 0.001831518835388124
step: 130, loss: 1.5612318748026155e-05
step: 140, loss: 1.2770162356900983e-05
step: 150, loss: 5.9991896705469117e-05
step: 160, loss: 9.325166320195422e-05
step: 170, loss: 0.06666726619005203
step: 180, loss: 2.871170363505371e-05
step: 190, loss: 0.0001178011589217931
step: 200, loss: 0.0007493176963180304
step: 210, loss: 0.0003240747028030455
step: 220, loss: 2.093878538289573e-05
step: 230, loss: 0.03249524161219597
step: 240, loss: 4.464695302885957e-05
step: 250, loss: 1.922585033753421e-05
step: 260, loss: 0.0002621082530822605
step: 270, loss: 0.21792669594287872
step: 280, loss: 0.0007042131619527936
step: 290, loss: 0.0012547958176583052
step: 300, loss: 0.0005885817809030414
step: 310, loss: 0.0005057977396063507
step: 320, loss: 0.0001240584097104147
step: 330, loss: 0.009229580871760845
step: 340, loss: 0.00046144952648319304
step: 350, loss: 0.0009602708742022514
step: 360, loss: 0.00013190349272917956
step: 370, loss: 0.00010124263644684106
step: 380, loss: 3.1554394809063524e-05
step: 390, loss: 0.000322275678627193
step: 400, loss: 1.8763257685350254e-05
step: 410, loss: 0.00016748261987231672
step: 420, loss: 4.867478128289804e-05
step: 430, loss: 4.5972854422871023e-05
step: 440, loss: 0.00978097878396511
step: 450, loss: 0.0012196892639622092
step: 460, loss: 4.1084571421379223e-05
step: 470, loss: 0.0001428856048732996
step: 480, loss: 0.00038943474646657705
step: 490, loss: 0.0001861877244664356
step: 500, loss: 2.178803515562322e-05
step: 510, loss: 4.272329533705488e-05
step: 520, loss: 2.7033103833673522e-05
step: 530, loss: 3.9202477637445554e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 15: dev_f1=0.937037037037037, f1=0.9389454209065681,
best_f1=0.9389454209065681
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
5742it [00:23, 245.38it/s]
load_f1 = 0.9358736059479555
real_f1 = 0.9330855018587362
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:17, 247.64it/s]

```

2 BASELINE

2.1 BASELINE STRUCTURED

2.1.1 Beer - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Structured/Beer \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Beer.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/Beer \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic
```

Processing user overrides (additional kwargs that are not None)...

```

After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8611232042312622
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.27184466019417475, f1=0.2772277227722772,
best_f1=0.2772277227722772
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.37596943974494934
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.41025641025641024, f1=0.32, best_f1=0.32
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```



```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3166946470737457
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.3714285714285714, f1=0.30434782608695654, best_f1=0.32
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.34267106652259827
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.43902439024390244, f1=0.39999999999999997,
best_f1=0.39999999999999997
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.22138170897960663
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.46511627906976755, f1=0.358974358974359,
best_f1=0.358974358974359
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3011865019798279
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 6: dev_f1=0.5625000000000001, f1=0.5294117647058824,
best_f1=0.5294117647058824
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.22254729270935059
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.4666666666666667, f1=0.3448275862068965,
best_f1=0.5294117647058824
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2882792353630066
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.5142857142857143, f1=0.3243243243243243,
best_f1=0.5294117647058824
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.20893129706382751
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.5625000000000001, f1=0.3333333333333333,
best_f1=0.5294117647058824
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.15998058021068573
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 10: dev_f1=0.5806451612903226, f1=0.5294117647058824,
best_f1=0.5294117647058824
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.134042426943779
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.5806451612903226, f1=0.5517241379310344,
best_f1=0.5294117647058824
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.16547316312789917
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 12: dev_f1=0.6086956521739131, f1=0.56, best_f1=0.56
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.118023581802845
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.6000000000000001, f1=0.6000000000000001, best_f1=0.56
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.147380068898201
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.6086956521739131, f1=0.5, best_f1=0.56
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2561447322368622
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.6086956521739131, f1=0.5, best_f1=0.56
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

```

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.

Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

91it [00:00, 134537.07it/s]

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

load_f1 = 0.6086956521739131

real_f1 = 0.5833333333333334

0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-

packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

4398it [00:11, 372.24it/s]

2.1.2 DBLP-ACM - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Structured/DBLP-ACM \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-ACM.pt

# Step 3: run the matcher
```

```
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
--task Structured/DBLP-ACM \
--input_path input/input_small.jsonl \
--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:

UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order: `optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

```
"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
```

```
step: 0, loss: 0.8118119835853577
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0

```

step: 10, loss: 0.4697166979312897
step: 20, loss: 0.5667679905891418
step: 30, loss: 0.40378037095069885
step: 40, loss: 0.18330073356628418
step: 50, loss: 0.238571435213089
step: 60, loss: 0.025980429723858833
step: 70, loss: 0.12636221945285797
step: 80, loss: 0.17620675265789032
step: 90, loss: 0.015462120063602924
step: 100, loss: 0.06107545271515846
step: 110, loss: 0.03870132565498352
step: 120, loss: 0.07447116822004318
step: 130, loss: 0.008816475979983807
step: 140, loss: 0.10836967080831528
step: 150, loss: 0.09714684635400772
step: 160, loss: 0.20285990834236145
step: 170, loss: 0.057421449571847916
step: 180, loss: 0.009391821920871735
step: 190, loss: 0.024332640692591667
step: 200, loss: 0.021348770707845688
step: 210, loss: 0.04157256707549095
step: 220, loss: 0.039492372423410416
step: 230, loss: 0.021568801254034042
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9797297297297298, f1=0.9774266365688488,
best_f1=0.9774266365688488
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.046713802963495255
step: 10, loss: 0.002201683120802045
step: 20, loss: 0.013751057907938957
step: 30, loss: 0.0011509229661896825
step: 40, loss: 0.0034982392098754644
step: 50, loss: 0.18457268178462982
step: 60, loss: 0.14381808042526245
step: 70, loss: 0.01270353328436613
step: 80, loss: 0.005079244263470173
step: 90, loss: 0.004734032321721315
step: 100, loss: 0.01704733446240425
step: 110, loss: 0.02532271482050419
step: 120, loss: 0.0016492248978465796

```

```

step: 130, loss: 0.000874209392350167
step: 140, loss: 0.29745879769325256
step: 150, loss: 0.034819308668375015
step: 160, loss: 0.11023398488759995
step: 170, loss: 0.03233170509338379
step: 180, loss: 0.009160608053207397
step: 190, loss: 0.10078167170286179
step: 200, loss: 0.001840149168856442
step: 210, loss: 0.03318643569946289
step: 220, loss: 0.0009339096723124385
step: 230, loss: 0.022107698023319244
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9831649831649831, f1=0.9742441209406495,
best_f1=0.9742441209406495
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04991685599088669
step: 10, loss: 0.0018974666018038988
step: 20, loss: 0.05514473840594292
step: 30, loss: 0.006547107361257076
step: 40, loss: 0.026726553216576576
step: 50, loss: 0.002163055120036006
step: 60, loss: 0.01860598661005497
step: 70, loss: 0.0019530976423993707
step: 80, loss: 0.0016971805598586798
step: 90, loss: 0.0033545196056365967
step: 100, loss: 0.0006844931049272418
step: 110, loss: 0.0018447430338710546
step: 120, loss: 0.0011596252443268895
step: 130, loss: 0.0026038270443677902
step: 140, loss: 0.0011743801878765225
step: 150, loss: 0.01332580391317606
step: 160, loss: 0.0021605202928185463
step: 170, loss: 0.00926517229527235
step: 180, loss: 0.0014005423290655017
step: 190, loss: 0.00375215127132833
step: 200, loss: 0.006009343080222607
step: 210, loss: 0.030974779278039932
step: 220, loss: 0.0014244221383705735
step: 230, loss: 0.021168077364563942
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.9713656387665198, f1=0.9723756906077348,
best_f1=0.9742441209406495
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002634374424815178
step: 10, loss: 0.0012379824183881283
step: 20, loss: 0.000572964025195688
step: 30, loss: 0.0007580366218462586
step: 40, loss: 0.003594567533582449
step: 50, loss: 0.008461609482765198
step: 60, loss: 0.00044958406942896545
step: 70, loss: 0.002047628862783313
step: 80, loss: 0.03802134841680527
step: 90, loss: 0.0026207275222986937
step: 100, loss: 0.08621973544359207
step: 110, loss: 0.01608498953282833
step: 120, loss: 0.048331767320632935
step: 130, loss: 0.0008260446484200656
step: 140, loss: 0.0009071767563000321
step: 150, loss: 0.0006638385821133852
step: 160, loss: 0.0003773655917029828
step: 170, loss: 0.013315041549503803
step: 180, loss: 0.07527877390384674
step: 190, loss: 0.003112387377768755
step: 200, loss: 0.0007444339571520686
step: 210, loss: 0.00018395765800960362
step: 220, loss: 0.0006229378050193191
step: 230, loss: 0.0009280262165702879
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.9841628959276018, f1=0.9829738933030647,
best_f1=0.9829738933030647
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003469160001259297
step: 10, loss: 0.03839683532714844
step: 20, loss: 0.00026897992938756943
step: 30, loss: 0.00014682360051665455
step: 40, loss: 0.001131246448494494

```



```

step: 50, loss: 0.00026633517700247467
step: 60, loss: 0.0002649743400979787
step: 70, loss: 0.00032150340848602355
step: 80, loss: 0.0005251215770840645
step: 90, loss: 0.011982516385614872
step: 100, loss: 0.003173480974510312
step: 110, loss: 0.002378721721470356
step: 120, loss: 0.043585069477558136
step: 130, loss: 0.0916367769241333
step: 140, loss: 0.00022337122936733067
step: 150, loss: 0.00015379491378553212
step: 160, loss: 0.03933529928326607
step: 170, loss: 0.03544756397604942
step: 180, loss: 0.0003847135812975466
step: 190, loss: 0.000492480059619993
step: 200, loss: 0.006949391681700945
step: 210, loss: 0.0001735167024889961
step: 220, loss: 0.00018980857566930354
step: 230, loss: 0.024169618263840675
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.9753914988814317, f1=0.9729119638826186,
best_f1=0.9829738933030647
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00028255642973817885
step: 10, loss: 0.08820688724517822
step: 20, loss: 0.003733938094228506
step: 30, loss: 0.0001599401730345562
step: 40, loss: 0.00045148737262934446
step: 50, loss: 0.1386571228504181
step: 60, loss: 0.013636463321745396
step: 70, loss: 0.021932849660515785
step: 80, loss: 0.013063213787972927
step: 90, loss: 0.002822232898324728
step: 100, loss: 0.00036374249611981213
step: 110, loss: 0.04692785069346428
step: 120, loss: 0.00016815029084682465
step: 130, loss: 0.00464213639497757
step: 140, loss: 0.0005098931142129004
step: 150, loss: 0.00032252786331810057
step: 160, loss: 0.011042093858122826
step: 170, loss: 0.0017107073217630386
step: 180, loss: 0.0005364476819522679
step: 190, loss: 0.00048106114263646305
step: 200, loss: 0.010629727505147457

```

```

step: 210, loss: 0.00039029892650432885
step: 220, loss: 0.0005911035696044564
step: 230, loss: 0.0012548798695206642
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 6: dev_f1=0.9853768278965129, f1=0.976324689966178,
best_f1=0.976324689966178
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04976997897028923
step: 10, loss: 0.00012456045078579336
step: 20, loss: 0.00036853691563010216
step: 30, loss: 0.00014675861166324466
step: 40, loss: 0.003308910643681884
step: 50, loss: 8.653357508592308e-05
step: 60, loss: 0.031031092628836632
step: 70, loss: 0.00012266659177839756
step: 80, loss: 0.00012990915274713188
step: 90, loss: 0.00041170904296450317
step: 100, loss: 0.00040930751129053533
step: 110, loss: 0.001646677148528397
step: 120, loss: 0.0009157621534541249
step: 130, loss: 0.0009348579915240407
step: 140, loss: 0.00016407971270382404
step: 150, loss: 0.09407363086938858
step: 160, loss: 0.0007098203059285879
step: 170, loss: 0.003640771145001054
step: 180, loss: 0.0032468244899064302
step: 190, loss: 0.02473127469420433
step: 200, loss: 0.00042020805994980037
step: 210, loss: 0.00019048279500566423
step: 220, loss: 0.00029565926524810493
step: 230, loss: 0.02564171515405178
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9819413092550789, f1=0.9739524348810873,
best_f1=0.976324689966178
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0209586713463068

```

```

step: 10, loss: 0.019731923937797546
step: 20, loss: 0.000279168103588745
step: 30, loss: 0.00019150492153130472
step: 40, loss: 0.00014958521933294833
step: 50, loss: 0.002856423379853368
step: 60, loss: 0.00022902201453689486
step: 70, loss: 0.00041412434075027704
step: 80, loss: 0.0015359976096078753
step: 90, loss: 0.00019430817337706685
step: 100, loss: 0.00017227848002221435
step: 110, loss: 0.00028000352904200554
step: 120, loss: 0.00016492436407133937
step: 130, loss: 0.0003376805398147553
step: 140, loss: 0.0006744720158167183
step: 150, loss: 9.870850044535473e-05
step: 160, loss: 0.00013853947166353464
step: 170, loss: 0.0001168043672805652
step: 180, loss: 0.0008471636101603508
step: 190, loss: 0.0009773875353857875
step: 200, loss: 0.014262838289141655
step: 210, loss: 7.147724681999534e-05
step: 220, loss: 0.00014029006706550717
step: 230, loss: 0.01870589330792427
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9853438556933484, f1=0.9774266365688488,
best_f1=0.976324689966178
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005011469474993646
step: 10, loss: 7.176099461503327e-05
step: 20, loss: 0.009361444041132927
step: 30, loss: 0.00017846653645392507
step: 40, loss: 5.722673449781723e-05
step: 50, loss: 7.49774553696625e-05
step: 60, loss: 0.00015762328985147178
step: 70, loss: 0.0001763649779604748
step: 80, loss: 0.0399799682199955
step: 90, loss: 0.0009881247533485293
step: 100, loss: 0.00010669746552594006
step: 110, loss: 7.088822167133912e-05
step: 120, loss: 0.03740609064698219
step: 130, loss: 0.00011433804320404306
step: 140, loss: 0.00011286672088317573
step: 150, loss: 8.741165220271796e-05
step: 160, loss: 0.00014344202645588666

```

```

step: 170, loss: 4.651058043236844e-05
step: 180, loss: 7.699986599618569e-05
step: 190, loss: 6.188551196828485e-05
step: 200, loss: 8.4413040895015e-05
step: 210, loss: 0.0017663772450760007
step: 220, loss: 0.05376075953245163
step: 230, loss: 0.0026580700650811195
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9796839729119639, f1=0.9751693002257337,
best_f1=0.976324689966178
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00016727155889384449
step: 10, loss: 0.00014510322944261134
step: 20, loss: 5.7619050494395196e-05
step: 30, loss: 8.660791354486719e-05
step: 40, loss: 0.0001320341689279303
step: 50, loss: 0.013760999776422977
step: 60, loss: 0.02893046662211418
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 70, loss: 0.00026438533677719533
step: 80, loss: 0.00010133330943062901
step: 90, loss: 7.256709068315104e-05
step: 100, loss: 0.0002562762820161879
step: 110, loss: 7.906325481599197e-05
step: 120, loss: 0.01949990540742874
step: 130, loss: 6.145602674223483e-05
step: 140, loss: 0.002423357218503952
step: 150, loss: 5.8785117289517075e-05
step: 160, loss: 0.00015620487101841718
step: 170, loss: 0.00031777567346580327
step: 180, loss: 0.000511817866936326
step: 190, loss: 8.201169111998752e-05
step: 200, loss: 5.241832695901394e-05
step: 210, loss: 5.095823871670291e-05
step: 220, loss: 0.0005074910586699843
step: 230, loss: 8.43793386593461e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 10: dev_f1=0.988814317673378, f1=0.9798657718120806,

```

```

best_f1=0.9798657718120806
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.6285465689143166e-05
step: 10, loss: 0.0003018011339008808
step: 20, loss: 2.9242613891256042e-05
step: 30, loss: 5.896336006117053e-05
step: 40, loss: 0.0008726929663680494
step: 50, loss: 0.00022610744053963572
step: 60, loss: 7.513164018746465e-05
step: 70, loss: 8.154923125403002e-05
step: 80, loss: 0.00010711799404816702
step: 90, loss: 0.00011130523489555344
step: 100, loss: 0.0015445761382579803
step: 110, loss: 5.57777239009738e-05
step: 120, loss: 6.298055814113468e-05
step: 130, loss: 7.964768883539364e-05
step: 140, loss: 8.548985351808369e-05
step: 150, loss: 7.107884448487312e-05
step: 160, loss: 8.38816340547055e-05
step: 170, loss: 0.003736883634701371
step: 180, loss: 0.00042627041693776846
step: 190, loss: 0.00010333150567021221
step: 200, loss: 6.444210157496855e-05
step: 210, loss: 4.744208126794547e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 220, loss: 5.7155004469677806e-05
step: 230, loss: 0.012899388559162617
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9887133182844244, f1=0.9796839729119639,
best_f1=0.9798657718120806
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001281530479900539
step: 10, loss: 0.00013128490536473691
step: 20, loss: 6.521498289657757e-05
step: 30, loss: 9.25003841985017e-05
step: 40, loss: 3.5444245440885425e-05
step: 50, loss: 4.718379204859957e-05
step: 60, loss: 0.011786727234721184
step: 70, loss: 0.00016417789447586983
step: 80, loss: 0.00013236847007647157
step: 90, loss: 5.312502253218554e-05
step: 100, loss: 0.00014429025759454817

```

```

step: 110, loss: 0.00031289030448533595
step: 120, loss: 9.104644414037466e-05
step: 130, loss: 9.970808605430648e-05
step: 140, loss: 4.3471100070746616e-05
step: 150, loss: 2.8419235604815185e-05
step: 160, loss: 0.029684970155358315
step: 170, loss: 0.00013929783017374575
step: 180, loss: 3.777991150855087e-05
step: 190, loss: 6.183476943988353e-05
step: 200, loss: 0.006724707316607237
step: 210, loss: 5.378005516831763e-05
step: 220, loss: 0.025246987119317055
step: 230, loss: 9.200075146509334e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9875706214689265, f1=0.9829738933030647,
best_f1=0.9798657718120806
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00014683314657304436
step: 10, loss: 9.722260438138619e-05
step: 20, loss: 8.88333743205294e-05
step: 30, loss: 0.025957223027944565
step: 40, loss: 0.00010569130245130509
step: 50, loss: 5.39720531378407e-05
step: 60, loss: 0.0003374095249455422
step: 70, loss: 4.571029057842679e-05
step: 80, loss: 4.057027035742067e-05
step: 90, loss: 5.3400875913212076e-05
step: 100, loss: 6.76454437780194e-05
step: 110, loss: 9.15165655896999e-05
step: 120, loss: 5.031801265431568e-05
step: 130, loss: 0.0017103059217333794
step: 140, loss: 0.001955739688128233
step: 150, loss: 0.018232837319374084
step: 160, loss: 3.809650297625922e-05
step: 170, loss: 5.1822051318595186e-05
step: 180, loss: 0.00011420142982387915
step: 190, loss: 3.0050810892134905e-05
step: 200, loss: 0.00021295958140399307
step: 210, loss: 5.965493619441986e-05
step: 220, loss: 2.874309939215891e-05
step: 230, loss: 6.60486621200107e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

epoch 13: dev_f1=0.9875706214689265, f1=0.9796380090497738,
best_f1=0.9798657718120806
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00014940049732103944
step: 10, loss: 2.5401377570233308e-05
step: 20, loss: 0.013295453041791916
step: 30, loss: 5.798689016955905e-05
step: 40, loss: 3.535745418048464e-05
step: 50, loss: 6.151991692604497e-05
step: 60, loss: 6.299898086581379e-05
step: 70, loss: 6.319709063973278e-05
step: 80, loss: 8.495946531184018e-05
step: 90, loss: 6.769037281628698e-05
step: 100, loss: 0.001880836091004312
step: 110, loss: 0.0005877861403860152
step: 120, loss: 0.00014205346815288067
step: 130, loss: 6.145572842797264e-05
step: 140, loss: 0.00010902028589043766
step: 150, loss: 5.156213592272252e-05
step: 160, loss: 3.316432048450224e-05
step: 170, loss: 3.142136847600341e-05
step: 180, loss: 0.0004020800697617233
step: 190, loss: 0.001554942224174738
step: 200, loss: 3.0151311875670217e-05
step: 210, loss: 0.000381641264539212
step: 220, loss: 0.00010003277566283941
step: 230, loss: 1.9091807189397514e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9887387387387387, f1=0.9774266365688488,
best_f1=0.9798657718120806
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.2207553633488715e-05
step: 10, loss: 0.0033652978017926216
step: 20, loss: 6.038924402673729e-05
step: 30, loss: 0.00015826224989723414
step: 40, loss: 5.7789326092461124e-05
step: 50, loss: 0.00010400540486443788
step: 60, loss: 3.887322236550972e-05
step: 70, loss: 4.078987694811076e-05
step: 80, loss: 0.024976272135972977
step: 90, loss: 2.5152254238491878e-05
step: 100, loss: 0.0005349984276108444

```

```

step: 110, loss: 6.328653398668393e-05
step: 120, loss: 0.00011290697875665501
step: 130, loss: 0.0001278710988117382
step: 140, loss: 6.49902067380026e-05
step: 150, loss: 7.017984171397984e-05
step: 160, loss: 7.363181794062257e-05
step: 170, loss: 0.0001552690373500809
step: 180, loss: 3.9590344385942444e-05
step: 190, loss: 0.010835395194590092
step: 200, loss: 5.4173069656826556e-05
step: 210, loss: 0.02340671978890896
step: 220, loss: 0.0002624643675517291
step: 230, loss: 5.753300501964986e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9887387387387387, f1=0.9785794813979707,
best_f1=0.9798657718120806
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```



```

2473it [00:10, 226.08it/s]
load_f1 = 0.9899216125419933
real_f1 = 0.9876819708846584
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:18, 241.47it/s]

```

2.1.3 DBLP-GoogleScholar - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Structured/DBLP-GoogleScholar \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-GoogleScholar.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/DBLP-GoogleScholar \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16

```

```

patch_torch_functions : False
keep_batchnorm_fp32   : True
master_weights        : True
loss_scale            : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled               : True
opt_level             : 02
cast_model_type       : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32   : True
master_weights        : True
loss_scale            : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8050600290298462
step: 10, loss: 0.41064029932022095
step: 20, loss: 0.49189698696136475
step: 30, loss: 0.39829251170158386
step: 40, loss: 0.2808760106563568
step: 50, loss: 0.150045707821846
step: 60, loss: 0.12081505358219147
step: 70, loss: 0.06640080362558365
step: 80, loss: 0.11690974980592728
step: 90, loss: 0.29538074135780334
step: 100, loss: 0.21535682678222656
step: 110, loss: 0.11309757828712463
step: 120, loss: 0.05710943788290024
step: 130, loss: 0.02335411123931408
step: 140, loss: 0.06051667034626007
step: 150, loss: 0.01989438571035862
step: 160, loss: 0.1241719126701355
step: 170, loss: 0.16342793405056
step: 180, loss: 0.15860088169574738

```

```

step: 190, loss: 0.08379790931940079
step: 200, loss: 0.12684474885463715
step: 210, loss: 0.09470798820257187
step: 220, loss: 0.057677946984767914
step: 230, loss: 0.13533154129981995
step: 240, loss: 0.1277061104774475
step: 250, loss: 0.1047959178686142
step: 260, loss: 0.020926114171743393
step: 270, loss: 0.015748174861073494
step: 280, loss: 0.06281202286481857
step: 290, loss: 0.04671454057097435
step: 300, loss: 0.13331575691699982
step: 310, loss: 0.13760702311992645
step: 320, loss: 0.12483248859643936
step: 330, loss: 0.16890205442905426
step: 340, loss: 0.15048819780349731
step: 350, loss: 0.14699861407279968
step: 360, loss: 0.052784357219934464
step: 370, loss: 0.12130136787891388
step: 380, loss: 0.13977529108524323
step: 390, loss: 0.07755912840366364
step: 400, loss: 0.013000517152249813
step: 410, loss: 0.04413466900587082
step: 420, loss: 0.012592646293342113
step: 430, loss: 0.11319764703512192
step: 440, loss: 0.029484553262591362
step: 450, loss: 0.05162204056978226
step: 460, loss: 0.12792713940143585
step: 470, loss: 0.3576861321926117
step: 480, loss: 0.20750682055950165
step: 490, loss: 0.022463148459792137
step: 500, loss: 0.010159916244447231
step: 510, loss: 0.05902070179581642
step: 520, loss: 0.0489254966378212
step: 530, loss: 0.200956329703331
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9272311212814646, f1=0.920402561756633,
best_f1=0.920402561756633
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1531253159046173

```

step: 10, loss: 0.136042982339859
step: 20, loss: 0.12577320635318756
step: 30, loss: 0.013242091052234173
step: 40, loss: 0.011040658690035343
step: 50, loss: 0.131864532828331
step: 60, loss: 0.28257834911346436
step: 70, loss: 0.04960016906261444
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 80, loss: 0.011409568600356579
step: 90, loss: 0.03837992995977402
step: 100, loss: 0.5789880752563477
step: 110, loss: 0.031360428780317307
step: 120, loss: 0.05312294512987137
step: 130, loss: 0.07006429135799408
step: 140, loss: 0.014594272710382938
step: 150, loss: 0.026514997705817223
step: 160, loss: 0.004691916983574629
step: 170, loss: 0.07358069717884064
step: 180, loss: 0.034076713025569916
step: 190, loss: 0.0821913480758667
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 200, loss: 0.013712975196540356
step: 210, loss: 0.025583676993846893
step: 220, loss: 0.0053308019414544106
step: 230, loss: 0.03624647483229637
step: 240, loss: 0.12350794672966003
step: 250, loss: 0.03141484782099724
step: 260, loss: 0.0325472429394722
step: 270, loss: 0.16887550055980682
step: 280, loss: 0.30354201793670654
step: 290, loss: 0.13134649395942688
step: 300, loss: 0.030183356255292892
step: 310, loss: 0.08424274623394012
step: 320, loss: 0.06962572783231735
step: 330, loss: 0.02863023616373539
step: 340, loss: 0.008771057240664959
step: 350, loss: 0.08083093166351318
step: 360, loss: 0.03377220034599304
step: 370, loss: 0.011799128726124763
step: 380, loss: 0.07513955980539322
step: 390, loss: 0.03631438687443733
step: 400, loss: 0.042784497141838074
step: 410, loss: 0.0002395008923485875
step: 420, loss: 0.04253527894616127
step: 430, loss: 0.03362058475613594
step: 440, loss: 0.019104093313217163
step: 450, loss: 0.015530393458902836
step: 460, loss: 0.2241394966840744

```

step: 470, loss: 0.041747692972421646
step: 480, loss: 0.5070406794548035
step: 490, loss: 0.06808310747146606
step: 500, loss: 0.019258201122283936
step: 510, loss: 0.050042085349559784
step: 520, loss: 0.2125670462846756
step: 530, loss: 0.1194465160369873
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9324074074074072, f1=0.9341372912801483,
best_f1=0.9341372912801483
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01057699415832758
step: 10, loss: 0.07398611307144165
step: 20, loss: 0.08312147110700607
step: 30, loss: 0.1943342536687851
step: 40, loss: 0.0035901269875466824
step: 50, loss: 0.014377507381141186
step: 60, loss: 0.013129189610481262
step: 70, loss: 0.05317959561944008
step: 80, loss: 0.007568359375
step: 90, loss: 0.007932798936963081
step: 100, loss: 0.002343863481655717
step: 110, loss: 0.01940334588289261
step: 120, loss: 0.012759683653712273
step: 130, loss: 0.013629479333758354
step: 140, loss: 0.1383545994758606
step: 150, loss: 0.003277151146903634
step: 160, loss: 0.006320074200630188
step: 170, loss: 0.00319411838427186
step: 180, loss: 0.001409611664712429
step: 190, loss: 0.006233959458768368
step: 200, loss: 0.018046509474515915
step: 210, loss: 0.038033049553632736
step: 220, loss: 0.04148603975772858
step: 230, loss: 0.015307551249861717
step: 240, loss: 0.027022933587431908
step: 250, loss: 0.007256143260747194
step: 260, loss: 0.0013037372846156359
step: 270, loss: 0.0038857641629874706
step: 280, loss: 0.047571249306201935

```

```

step: 290, loss: 0.08649677038192749
step: 300, loss: 0.06786902993917465
step: 310, loss: 0.06244078651070595
step: 320, loss: 0.16390478610992432
step: 330, loss: 0.004101656377315521
step: 340, loss: 0.001752606243826449
step: 350, loss: 0.033192284405231476
step: 360, loss: 0.07907895743846893
step: 370, loss: 0.01996777206659317
step: 380, loss: 0.01675204001367092
step: 390, loss: 0.03604529798030853
step: 400, loss: 0.035909898579120636
step: 410, loss: 0.005679022986441851
step: 420, loss: 0.07008638978004456
step: 430, loss: 0.006626718677580357
step: 440, loss: 0.005163806024938822
step: 450, loss: 0.16059346497058868
step: 460, loss: 0.04430490732192993
step: 470, loss: 0.007631026208400726
step: 480, loss: 0.018658967688679695
step: 490, loss: 0.015841199085116386
step: 500, loss: 0.12419890612363815
step: 510, loss: 0.005462686996906996
step: 520, loss: 0.03190217167139053
step: 530, loss: 0.05180247873067856
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.9316596931659693, f1=0.9304267161410019,
best_f1=0.9341372912801483
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004041858483105898
step: 10, loss: 0.0060765380039811134
step: 20, loss: 0.06591907143592834
step: 30, loss: 0.060666780918836594
step: 40, loss: 0.005896066315472126
step: 50, loss: 0.027293451130390167
step: 60, loss: 0.004360392689704895
step: 70, loss: 0.0023070729803293943
step: 80, loss: 0.01800530217587948
step: 90, loss: 0.0031964436639100313
step: 100, loss: 0.0033663311041891575
step: 110, loss: 0.0008826546836644411
step: 120, loss: 0.0010187687585130334
step: 130, loss: 0.031598228961229324
step: 140, loss: 0.0020678096916526556

```

```

step: 150, loss: 0.0008544893353246152
step: 160, loss: 0.19047099351882935
step: 170, loss: 0.0042761387303471565
step: 180, loss: 0.0034223657567054033
step: 190, loss: 0.005659158807247877
step: 200, loss: 0.002765327226370573
step: 210, loss: 0.012052271515130997
step: 220, loss: 0.011773142963647842
step: 230, loss: 0.04250537231564522
step: 240, loss: 0.0012981615727767348
step: 250, loss: 0.00014781815116293728
step: 260, loss: 0.21191084384918213
step: 270, loss: 0.038077812641859055
step: 280, loss: 0.014004182070493698
step: 290, loss: 0.12168630957603455
step: 300, loss: 0.0012342905392870307
step: 310, loss: 0.03128466382622719
step: 320, loss: 0.18387199938297272
step: 330, loss: 0.025367485359311104
step: 340, loss: 0.04178521782159805
step: 350, loss: 0.009075275622308254
step: 360, loss: 0.03530585765838623
step: 370, loss: 0.04259341582655907
step: 380, loss: 0.02499624527990818
step: 390, loss: 0.04089156910777092
step: 400, loss: 0.003953117877244949
step: 410, loss: 0.03830854594707489
step: 420, loss: 0.011291713453829288
step: 430, loss: 0.0061098006553947926
step: 440, loss: 0.051496319472789764
step: 450, loss: 0.006226047873497009
step: 460, loss: 0.003383810166269541
step: 470, loss: 0.019093744456768036
step: 480, loss: 0.0012306743301451206
step: 490, loss: 0.012644925154745579
step: 500, loss: 0.00496671861037612
step: 510, loss: 0.06373002380132675
step: 520, loss: 0.005336327012628317
step: 530, loss: 0.0050796810537576675
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.9357967667436491, f1=0.9291338582677167,
best_f1=0.9291338582677167

```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:  
An input tensor was not cuda.
```

```
    warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.0012145000509917736  
step: 10, loss: 0.0020277095027267933  
step: 20, loss: 0.025770802050828934  
step: 30, loss: 0.00037753695505671203  
step: 40, loss: 0.027702338993549347  
step: 50, loss: 0.0010576439090073109  
step: 60, loss: 0.00025650972384028137  
step: 70, loss: 0.00017350474081467837  
step: 80, loss: 0.0018883139127865434  
step: 90, loss: 0.08385467529296875  
step: 100, loss: 0.0009598183678463101  
step: 110, loss: 0.0037132566794753075  
step: 120, loss: 0.0034263799898326397  
step: 130, loss: 0.008610283955931664  
step: 140, loss: 0.002285764319822192  
step: 150, loss: 0.015300987288355827  
step: 160, loss: 0.14446298778057098  
step: 170, loss: 0.003199593396857381  
step: 180, loss: 0.006817881017923355  
step: 190, loss: 0.007052707951515913  
step: 200, loss: 0.15112201869487762  
step: 210, loss: 0.0012011440703645349  
step: 220, loss: 0.0015063588507473469  
step: 230, loss: 0.002607228234410286  
step: 240, loss: 0.002321587409824133  
step: 250, loss: 0.001404867391102016  
step: 260, loss: 0.0024699310306459665  
step: 270, loss: 0.0833219513297081  
step: 280, loss: 0.037425216287374496  
step: 290, loss: 0.00031312424107454717  
step: 300, loss: 0.0217438954859972  
step: 310, loss: 0.0016402872279286385  
step: 320, loss: 0.015274642035365105  
step: 330, loss: 0.0005857351352460682  
step: 340, loss: 0.02193499729037285  
step: 350, loss: 0.0015936987474560738  
step: 360, loss: 0.03633641451597214  
step: 370, loss: 0.01956668123602867  
step: 380, loss: 0.005473915487527847  
step: 390, loss: 0.001151342294178903  
step: 400, loss: 0.14056633412837982  
step: 410, loss: 0.0007333853282034397  
step: 420, loss: 0.006669835187494755  
step: 430, loss: 0.0015232297591865063  
step: 440, loss: 0.004394617397338152
```



```

step: 450, loss: 0.0005809312569908798
step: 460, loss: 0.0026203314773738384
step: 470, loss: 0.0020986658055335283
step: 480, loss: 0.0009554260177537799
step: 490, loss: 0.010283202864229679
step: 500, loss: 0.012701011262834072
step: 510, loss: 0.052283041179180145
step: 520, loss: 0.001541375881060958
step: 530, loss: 0.003412303514778614
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.9400278940027894, f1=0.9295380307979467,
best_f1=0.9295380307979467
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0015045866603031754
step: 10, loss: 0.015268324874341488
step: 20, loss: 0.017108891159296036
step: 30, loss: 0.008248134516179562
step: 40, loss: 0.005502194166183472
step: 50, loss: 0.04128185287117958
step: 60, loss: 0.0647948607802391
step: 70, loss: 0.004043933004140854
step: 80, loss: 0.0015254892641678452
step: 90, loss: 0.0048650833778083324
step: 100, loss: 0.17357119917869568
step: 110, loss: 0.1527635157108307
step: 120, loss: 0.03496871888637543
step: 130, loss: 0.0006520209135487676
step: 140, loss: 0.001673996914178133
step: 150, loss: 0.0015661605866625905
step: 160, loss: 0.00030179580789990723
step: 170, loss: 0.0003161344793625176
step: 180, loss: 0.0002579391875769943
step: 190, loss: 0.003501133294776082
step: 200, loss: 0.0013276637764647603
step: 210, loss: 0.0005133473896421492
step: 220, loss: 0.0013020525220781565
step: 230, loss: 0.0008905089343897998
step: 240, loss: 0.000501251604873687
step: 250, loss: 0.00020134825899731368
step: 260, loss: 0.0002592840464785695

```

```
step: 270, loss: 0.0029448694549500942
step: 280, loss: 0.004269816912710667
step: 290, loss: 0.0007360241143032908
step: 300, loss: 0.0005483436398208141
step: 310, loss: 0.0006614249432459474
step: 320, loss: 0.003098058048635721
step: 330, loss: 0.014758145436644554
step: 340, loss: 0.0005698254099115729
step: 350, loss: 0.12649786472320557
step: 360, loss: 0.008211144246160984
step: 370, loss: 0.004387832712382078
step: 380, loss: 0.0026964624412357807
step: 390, loss: 0.011750392615795135
step: 400, loss: 0.0003760427061934024
step: 410, loss: 0.0005335891619324684
step: 420, loss: 0.11070737987756729
step: 430, loss: 0.0009826843161135912
step: 440, loss: 0.008842779323458672
step: 450, loss: 0.08585266768932343
step: 460, loss: 0.008215060457587242
step: 470, loss: 0.13539178669452667
step: 480, loss: 0.017147917300462723
step: 490, loss: 0.0006364636356011033
step: 500, loss: 0.006002564448863268
step: 510, loss: 0.0003464336332399398
step: 520, loss: 0.0010616622166708112
step: 530, loss: 0.0007331466185860336
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.9304713019132057, f1=0.9191210846189808,
best_f1=0.9295380307979467
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00136876804754138
step: 10, loss: 0.019273756071925163
step: 20, loss: 0.03419272601604462
step: 30, loss: 0.0019307589391246438
step: 40, loss: 0.0004008487449027598
step: 50, loss: 0.0009038049611262977
step: 60, loss: 0.0005507150199264288
step: 70, loss: 0.05477813258767128
step: 80, loss: 0.008490226231515408
step: 90, loss: 0.00032374492730014026
step: 100, loss: 0.0025672006886452436
step: 110, loss: 0.00017977935203816742
step: 120, loss: 0.0011793577577918768
```

```

step: 130, loss: 0.00012518376752268523
step: 140, loss: 0.0012091826647520065
step: 150, loss: 9.971673716790974e-05
step: 160, loss: 0.00040018456638790667
step: 170, loss: 0.0004348666116129607
step: 180, loss: 0.00013676792150363326
step: 190, loss: 0.00034980798955075443
step: 200, loss: 0.0030843629501760006
step: 210, loss: 6.658105849055573e-05
step: 220, loss: 8.06438984000124e-05
step: 230, loss: 0.00034820070140995085
step: 240, loss: 0.00040661400998942554
step: 250, loss: 0.0012585449730977416
step: 260, loss: 0.0002647456421982497
step: 270, loss: 0.0015407700557261705
step: 280, loss: 0.04199851676821709
step: 290, loss: 0.00352029362693429
step: 300, loss: 0.002324402332305908
step: 310, loss: 0.0018754996126517653
step: 320, loss: 0.02658422291278839
step: 330, loss: 0.000189078418770805
step: 340, loss: 0.004297706298530102
step: 350, loss: 0.0004299394495319575
step: 360, loss: 0.0007935629109852016
step: 370, loss: 0.0009176833555102348
step: 380, loss: 0.005477468483150005
step: 390, loss: 0.00018805652507580817
step: 400, loss: 7.391376129817218e-05
step: 410, loss: 0.00017674139235168695
step: 420, loss: 0.00010790001397253945
step: 430, loss: 7.891887071309611e-05
step: 440, loss: 0.00020514809875749052
step: 450, loss: 0.0001242748403456062
step: 460, loss: 0.004803084768354893
step: 470, loss: 0.001006360282190144
step: 480, loss: 0.0020666555501520634
step: 490, loss: 0.0005625910125672817
step: 500, loss: 7.813550473656505e-05
step: 510, loss: 0.004241071175783873
step: 520, loss: 0.0002982076839543879
step: 530, loss: 0.0001846666564233601
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.937095282146161, f1=0.9262865090403337,
best_f1=0.9295380307979467
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0011812758166342974
step: 10, loss: 0.0026200313586741686
step: 20, loss: 0.013474510051310062
step: 30, loss: 0.0008296847227029502
step: 40, loss: 0.0007704311283305287
step: 50, loss: 0.0003420051361899823
step: 60, loss: 0.004662209656089544
step: 70, loss: 0.0005034453934058547
step: 80, loss: 0.00028470595134422183
step: 90, loss: 0.00037479682941921055
step: 100, loss: 0.0004421818011905998
step: 110, loss: 0.012691284529864788
step: 120, loss: 0.008826974779367447
step: 130, loss: 0.00016641570255160332
step: 140, loss: 5.93024305999279e-05
step: 150, loss: 0.00042721052886918187
step: 160, loss: 0.0001887109101517126
step: 170, loss: 0.0019571001175791025
step: 180, loss: 0.0005856394418515265
step: 190, loss: 0.002959203440696001
step: 200, loss: 0.0032462102826684713
step: 210, loss: 0.0001567705621710047
step: 220, loss: 0.0005510838818736374
step: 230, loss: 0.0004370566748548299
step: 240, loss: 0.00012807577149942517
step: 250, loss: 0.014776693657040596
step: 260, loss: 0.024412844330072403
step: 270, loss: 6.135144940344617e-05
step: 280, loss: 0.016887208446860313
step: 290, loss: 0.01597975566983223
step: 300, loss: 4.431389970704913e-05
step: 310, loss: 0.028607171028852463
step: 320, loss: 0.02505919709801674
step: 330, loss: 0.004484424367547035
step: 340, loss: 0.003209722461178899
step: 350, loss: 0.04446514695882797
step: 360, loss: 0.049608927220106125
step: 370, loss: 0.0003228327550459653
step: 380, loss: 0.0008331871940754354
step: 390, loss: 9.636974573368207e-05
step: 400, loss: 0.0460742749273777
step: 410, loss: 0.011332600377500057
step: 420, loss: 8.834899927023798e-05
step: 430, loss: 0.00019518767658155411
step: 440, loss: 0.00010367011418566108
step: 450, loss: 0.0003924200136680156
step: 460, loss: 0.0002691857807803899
```

```

step: 470, loss: 8.94934346433729e-05
step: 480, loss: 0.0002515507803764194
step: 490, loss: 0.00013727362966164947
step: 500, loss: 0.001155751058831811
step: 510, loss: 4.0362359868595377e-05
step: 520, loss: 0.00012630660785362124
step: 530, loss: 0.00012561686162371188
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 8: dev_f1=0.9401789919924635, f1=0.9268755935422602,
best_f1=0.9268755935422602
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00010680616833269596
step: 10, loss: 0.00041595145012252033
step: 20, loss: 0.0003421995206736028
step: 30, loss: 0.003225568449124694
step: 40, loss: 0.009279155172407627
step: 50, loss: 0.0004162652767263353
step: 60, loss: 0.0005534038064070046
step: 70, loss: 0.09281286597251892
step: 80, loss: 0.00036159498267807066
step: 90, loss: 0.030344117432832718
step: 100, loss: 0.0011473560007289052
step: 110, loss: 0.0015473837265744805
step: 120, loss: 0.00023038184735924006
step: 130, loss: 0.001231411937624216
step: 140, loss: 0.0032636078540235758
step: 150, loss: 0.00013708960614167154
step: 160, loss: 6.269930599955842e-05
step: 170, loss: 0.00011006673594238237
step: 180, loss: 0.0005310762790031731
step: 190, loss: 0.0009340718388557434
step: 200, loss: 0.00016809915541671216
step: 210, loss: 7.404525968013331e-05
step: 220, loss: 0.0004576544160954654
step: 230, loss: 7.651172927580774e-05
step: 240, loss: 0.1968850940465927
step: 250, loss: 0.002717986237257719
step: 260, loss: 0.00019280787091702223
step: 270, loss: 0.0006342710694298148
step: 280, loss: 9.70360342762433e-05

```

```

step: 290, loss: 8.799803617876023e-05
step: 300, loss: 0.00028427926008589566
step: 310, loss: 4.815025386051275e-05
step: 320, loss: 0.00021855383238289505
step: 330, loss: 0.00046770376502536237
step: 340, loss: 0.01730923354625702
step: 350, loss: 4.8410835006507114e-05
step: 360, loss: 0.02943095564842224
step: 370, loss: 0.00012718455400317907
step: 380, loss: 5.7833767641568556e-05
step: 390, loss: 0.0001411279517924413
step: 400, loss: 0.05743667110800743
step: 410, loss: 0.00030549956136383116
step: 420, loss: 0.0001687944313744083
step: 430, loss: 4.135978451813571e-05
step: 440, loss: 0.000196028413483873
step: 450, loss: 4.8492267524125054e-05
step: 460, loss: 6.090682290960103e-05
step: 470, loss: 3.880419535562396e-05
step: 480, loss: 0.0008236923022195697
step: 490, loss: 8.852085011312738e-05
step: 500, loss: 0.004621509462594986
step: 510, loss: 0.0013229487231001258
step: 520, loss: 0.0003933504631277174
step: 530, loss: 0.00017379847122356296
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9375879868606289, f1=0.9248826291079812,
best_f1=0.9268755935422602
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0013118069618940353
step: 10, loss: 2.219857560703531e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 20, loss: 0.0005790540017187595
step: 30, loss: 4.025327507406473e-05
step: 40, loss: 6.11219511483796e-05
step: 50, loss: 0.00011119770351797342
step: 60, loss: 0.0001676865213084966
step: 70, loss: 0.0001686617761151865
step: 80, loss: 6.193783337948844e-05
step: 90, loss: 0.00023484582197852433
step: 100, loss: 7.727937190793455e-05
step: 110, loss: 0.000125249513075687
step: 120, loss: 0.00010247723548673093
step: 130, loss: 8.617150888312608e-05

```

```

step: 140, loss: 7.846616063034162e-05
step: 150, loss: 0.0023763636127114296
step: 160, loss: 4.812023325939663e-05
step: 170, loss: 5.739870175602846e-05
step: 180, loss: 0.0015091905370354652
step: 190, loss: 0.0003426030743867159
step: 200, loss: 8.221478492487222e-05
step: 210, loss: 0.0005858403164893389
step: 220, loss: 6.506195495603606e-05
step: 230, loss: 0.0001917257613968104
step: 240, loss: 4.276839172234759e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 250, loss: 0.00039288646075874567
step: 260, loss: 0.002276506507769227
step: 270, loss: 0.00012697267811745405
step: 280, loss: 4.8976617108564824e-05
step: 290, loss: 6.816001405240968e-05
step: 300, loss: 0.0001607543817954138
step: 310, loss: 0.00029111222829669714
step: 320, loss: 0.0002066387824015692
step: 330, loss: 0.00013964693062007427
step: 340, loss: 0.0010334629332646728
step: 350, loss: 6.480458978330716e-05
step: 360, loss: 5.839582809130661e-05
step: 370, loss: 0.00010979536455124617
step: 380, loss: 0.00012225443788338453
step: 390, loss: 0.00040362804429605603
step: 400, loss: 0.011918751522898674
step: 410, loss: 7.264507439685985e-05
step: 420, loss: 0.0027585425414144993
step: 430, loss: 5.378744026529603e-05
step: 440, loss: 3.871555600198917e-05
step: 450, loss: 0.0004761120362672955
step: 460, loss: 7.475460733985528e-05
step: 470, loss: 3.490106973913498e-05
step: 480, loss: 9.987770317820832e-05
step: 490, loss: 0.04891984909772873
step: 500, loss: 0.0004515728505793959
step: 510, loss: 0.00019953076844103634
step: 520, loss: 0.0005797080229967833
step: 530, loss: 2.36400737776421e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)

```

```
epoch 10: dev_f1=0.9417840375586854, f1=0.9279321714554876,  
best_f1=0.9279321714554876  
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:  
An input tensor was not cuda.  
    warnings.warn("An input tensor was not cuda.")  
step: 0, loss: 7.562138489447534e-05  
step: 10, loss: 0.0003055354463867843  
step: 20, loss: 2.414303526165895e-05  
step: 30, loss: 0.0013331659138202667  
step: 40, loss: 0.0004307203635107726  
step: 50, loss: 0.0003560559416655451  
step: 60, loss: 4.904590605292469e-05  
step: 70, loss: 3.0266934118117206e-05  
step: 80, loss: 0.000290717143798247  
step: 90, loss: 0.00016051997954491526  
step: 100, loss: 0.000449770363047719  
step: 110, loss: 0.0009961610194295645  
step: 120, loss: 8.11387290013954e-05  
step: 130, loss: 3.418495543883182e-05  
step: 140, loss: 2.1583959096460603e-05  
step: 150, loss: 6.771518383175135e-05  
step: 160, loss: 0.00042283328366465867  
step: 170, loss: 0.0002233852428616956  
step: 180, loss: 9.266333654522896e-05  
step: 190, loss: 9.398970723850653e-05  
step: 200, loss: 0.0004496022593230009  
step: 210, loss: 9.299229714088142e-05  
step: 220, loss: 0.002923991298303008  
step: 230, loss: 0.1854637712240219  
step: 240, loss: 0.00028204810223542154  
step: 250, loss: 0.0009272117167711258  
step: 260, loss: 0.0002712332352530211  
step: 270, loss: 0.0012052023084834218  
step: 280, loss: 0.0001579955714987591  
step: 290, loss: 0.012698083184659481  
step: 300, loss: 0.04240294545888901  
step: 310, loss: 0.0002198332076659426  
step: 320, loss: 0.02493889443576336  
step: 330, loss: 5.510164191946387e-05  
step: 340, loss: 0.00016566291742492467  
step: 350, loss: 0.0018851181957870722  
step: 360, loss: 8.56333936098963e-05  
step: 370, loss: 2.7771076929639094e-05  
step: 380, loss: 7.847815868444741e-05  
step: 390, loss: 0.0012839477276429534  
step: 400, loss: 2.6418865672894754e-05  
step: 410, loss: 0.000434146320912987  
step: 420, loss: 0.0008033957565203309
```



```

step: 430, loss: 0.01980028674006462
step: 440, loss: 9.137346205534413e-05
step: 450, loss: 0.0003191889263689518
step: 460, loss: 0.004952464252710342
step: 470, loss: 0.0017217476852238178
step: 480, loss: 0.00010509350977372378
step: 490, loss: 0.0004998319200240076
step: 500, loss: 5.543197403312661e-05
step: 510, loss: 4.5474007492884994e-05
step: 520, loss: 8.344947127625346e-05
step: 530, loss: 6.311699689831585e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9364055299539169, f1=0.9320388349514563,
best_f1=0.9279321714554876
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003323840501252562
step: 10, loss: 0.0002824861148837954
step: 20, loss: 6.678208592347801e-05
step: 30, loss: 0.00012327761214692146
step: 40, loss: 0.00016497343312948942
step: 50, loss: 0.0007688052719458938
step: 60, loss: 0.0009511460666544735
step: 70, loss: 0.002316521480679512
step: 80, loss: 0.0002852145698852837
step: 90, loss: 0.00336826010607183
step: 100, loss: 0.00014766414824407548
step: 110, loss: 3.9105983887566254e-05
step: 120, loss: 0.00016529130516573787
step: 130, loss: 9.968047379516065e-05
step: 140, loss: 3.80964957003016e-05
step: 150, loss: 0.0001839887845562771
step: 160, loss: 4.769907536683604e-05
step: 170, loss: 0.00020190319628454745
step: 180, loss: 6.016336919856258e-05
step: 190, loss: 3.357695823069662e-05
step: 200, loss: 0.002973687369376421
step: 210, loss: 0.00027665012748911977
step: 220, loss: 4.5820630475645885e-05
step: 230, loss: 0.005839223507791758
step: 240, loss: 0.0001879088522400707
step: 250, loss: 3.360770278959535e-05
step: 260, loss: 0.0002849731536116451
step: 270, loss: 0.0002257231972180307
step: 280, loss: 4.950112270307727e-05

```

```

step: 290, loss: 0.000599766499362886
step: 300, loss: 0.00021931430092081428
step: 310, loss: 5.664022683049552e-05
step: 320, loss: 9.074692206922919e-05
step: 330, loss: 3.686738637043163e-05
step: 340, loss: 0.000521831214427948
step: 350, loss: 0.014886420220136642
step: 360, loss: 0.0003212690935470164
step: 370, loss: 4.5411099563352764e-05
step: 380, loss: 0.00022165104746818542
step: 390, loss: 4.354086559033021e-05
step: 400, loss: 0.001032872125506401
step: 410, loss: 0.00024112335813697428
step: 420, loss: 0.0006641848594881594
step: 430, loss: 0.00012148030509706587
step: 440, loss: 8.281806367449462e-05
step: 450, loss: 0.00013517533079721034
step: 460, loss: 0.003486282890662551
step: 470, loss: 3.214706157450564e-05
step: 480, loss: 8.369600982405245e-05
step: 490, loss: 0.0012959521263837814
step: 500, loss: 0.0005604312173090875
step: 510, loss: 0.003388266311958432
step: 520, loss: 0.019653787836432457
step: 530, loss: 4.211132545606233e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9394650398873767, f1=0.9292076887013595,
best_f1=0.9279321714554876
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.773618249804713e-05
step: 10, loss: 0.00035353907151147723
step: 20, loss: 0.026451831683516502
step: 30, loss: 7.947206904646009e-05
step: 40, loss: 2.7521558877197094e-05
step: 50, loss: 0.00028851881506852806
step: 60, loss: 7.10782187525183e-05
step: 70, loss: 2.5867531803669408e-05
step: 80, loss: 6.788827158743516e-05
step: 90, loss: 0.00014701587497256696
step: 100, loss: 2.3181375581771135e-05
step: 110, loss: 6.194121669977903e-05
step: 120, loss: 0.0001527815475128591
step: 130, loss: 0.0009179605985991657
step: 140, loss: 4.203775097266771e-05

```

```

step: 150, loss: 0.00023996926029212773
step: 160, loss: 3.2828411349328235e-05
step: 170, loss: 6.222131196409464e-05
step: 180, loss: 2.9238197384984232e-05
step: 190, loss: 0.00014824663230683655
step: 200, loss: 3.747552545974031e-05
step: 210, loss: 4.784801785717718e-05
step: 220, loss: 2.5420577003387734e-05
step: 230, loss: 0.0005392773309722543
step: 240, loss: 4.0227463614428416e-05
step: 250, loss: 0.04862380772829056
step: 260, loss: 9.554635471431538e-05
step: 270, loss: 3.1324529118137434e-05
step: 280, loss: 0.00010281433787895367
step: 290, loss: 5.681804032064974e-05
step: 300, loss: 3.165651287417859e-05
step: 310, loss: 0.00012216618051752448
step: 320, loss: 5.4797019402030855e-05
step: 330, loss: 5.744646114180796e-05
step: 340, loss: 2.7793699700850993e-05
step: 350, loss: 0.04148117080330849
step: 360, loss: 1.9073115254286677e-05
step: 370, loss: 3.13294876832515e-05
step: 380, loss: 0.0005260456237010658
step: 390, loss: 3.516827564453706e-05
step: 400, loss: 2.4984479750855826e-05
step: 410, loss: 4.930754585075192e-05
step: 420, loss: 6.36671029496938e-05
step: 430, loss: 0.003600197611376643
step: 440, loss: 0.00011096255184384063
step: 450, loss: 8.027713192859665e-05
step: 460, loss: 9.606728417566046e-05
step: 470, loss: 0.02916751615703106
step: 480, loss: 0.00033376316423527896
step: 490, loss: 0.019454045221209526
step: 500, loss: 8.732474816497415e-05
step: 510, loss: 0.00011871282185893506
step: 520, loss: 2.812827005982399e-05
step: 530, loss: 0.00020363258954603225
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9392265193370165, f1=0.9294605809128631,
best_f1=0.9279321714554876
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.0339084080187604e-05

```

step: 10, loss: 6.553380080731586e-05
step: 20, loss: 2.2407177311833948e-05
step: 30, loss: 8.048969903029501e-05
step: 40, loss: 9.577137097949162e-05
step: 50, loss: 4.1384235373698175e-05
step: 60, loss: 0.00031199550721794367
step: 70, loss: 4.456070018932223e-05
step: 80, loss: 0.001313935499638319
step: 90, loss: 1.85702938324539e-05
step: 100, loss: 7.982838724274188e-05
step: 110, loss: 3.533126073307358e-05
step: 120, loss: 2.0033745386172086e-05
step: 130, loss: 6.117181328590959e-05
step: 140, loss: 0.00958810281008482
step: 150, loss: 5.2433417295105755e-05
step: 160, loss: 0.00025671659386716783
step: 170, loss: 4.959735815646127e-05
step: 180, loss: 0.00013183706323616207
step: 190, loss: 0.00011772967991419137
step: 200, loss: 0.0005200615269131958
step: 210, loss: 8.633031393401325e-05
step: 220, loss: 5.813468305859715e-05
step: 230, loss: 3.4281569242011756e-05
step: 240, loss: 2.4429802579106763e-05
step: 250, loss: 3.0021232305443846e-05
step: 260, loss: 1.6674193830112927e-05
step: 270, loss: 0.0005054792854934931
step: 280, loss: 3.1673942430643365e-05
step: 290, loss: 4.178180097369477e-05
step: 300, loss: 2.8272659619688056e-05
step: 310, loss: 0.0002982396981678903
step: 320, loss: 4.4415657612262294e-05
step: 330, loss: 2.9606360840261914e-05
step: 340, loss: 3.542502599884756e-05
step: 350, loss: 4.938368874718435e-05
step: 360, loss: 0.0005043278215453029
step: 370, loss: 4.3006828491343185e-05
step: 380, loss: 2.2310210624709725e-05
step: 390, loss: 5.6783886975608766e-05
step: 400, loss: 3.350786937517114e-05
step: 410, loss: 2.2842998077976517e-05
step: 420, loss: 2.6649695428204723e-05
step: 430, loss: 2.8385240511852317e-05
step: 440, loss: 1.4200666555552743e-05
step: 450, loss: 3.308543819002807e-05
step: 460, loss: 2.930215487140231e-05
step: 470, loss: 7.424404611811042e-05
step: 480, loss: 2.8888287488371134e-05

```

step: 490, loss: 8.719359175302088e-05
step: 500, loss: 7.373550033662468e-05
step: 510, loss: 3.037441274500452e-05
step: 520, loss: 2.5163768441416323e-05
step: 530, loss: 0.00019298592815175653
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9392369288742346, f1=0.9307583608101743,
best_f1=0.9279321714554876
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 1.7843836758402176e-05
step: 10, loss: 0.0003078924783039838
step: 20, loss: 0.00016855461581144482
step: 30, loss: 8.050152973737568e-05
step: 40, loss: 2.4677887267898768e-05
step: 50, loss: 0.00030164900817908347
step: 60, loss: 1.8503089449950494e-05
step: 70, loss: 7.087168341968209e-05
step: 80, loss: 6.53000024612993e-05
step: 90, loss: 2.8362635930534452e-05
step: 100, loss: 2.285377740918193e-05
step: 110, loss: 0.018544821068644524
step: 120, loss: 6.151886191219091e-05
step: 130, loss: 0.0001217738026753068
step: 140, loss: 1.5262299712048844e-05
step: 150, loss: 4.9658581701805815e-05
step: 160, loss: 0.00013017715536989272
step: 170, loss: 0.00011415408516768366
step: 180, loss: 2.900714571296703e-05
step: 190, loss: 0.00019471452105790377
step: 200, loss: 4.425126462592743e-05
step: 210, loss: 2.5762969016795978e-05
step: 220, loss: 1.4629058568971232e-05
step: 230, loss: 9.998522727983072e-05
step: 240, loss: 3.7637200875906274e-05
step: 250, loss: 0.00015306888963095844
step: 260, loss: 0.00012128529488109052
step: 270, loss: 6.272904283832759e-05
step: 280, loss: 0.00020203963504172862
step: 290, loss: 0.00014754050062038004
step: 300, loss: 0.00148738082498312
step: 310, loss: 0.00010075055615743622
step: 320, loss: 0.0001006649254122749
step: 330, loss: 4.6898392611183226e-05
step: 340, loss: 8.137967233778909e-05

```

```

step: 350, loss: 3.262060636188835e-05
step: 360, loss: 0.0006753602647222579
step: 370, loss: 1.9203427655156702e-05
step: 380, loss: 2.1379037207225338e-05
step: 390, loss: 3.17547564918641e-05
step: 400, loss: 2.1445919628604315e-05
step: 410, loss: 0.00017189898062497377
step: 420, loss: 3.999183536507189e-05
step: 430, loss: 2.7130061425850727e-05
step: 440, loss: 8.73293392942287e-05
step: 450, loss: 0.01107131689786911
step: 460, loss: 1.4990388990554493e-05
step: 470, loss: 5.148869240656495e-05
step: 480, loss: 3.0303795938380063e-05
step: 490, loss: 0.0004611019685398787
step: 500, loss: 0.00010420572652947158
step: 510, loss: 2.906324152718298e-05
step: 520, loss: 0.00021648065012414008
step: 530, loss: 4.125606210436672e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9404536862003782, f1=0.9280983916745505,
best_f1=0.9279321714554876
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
5742it [00:23, 246.67it/s]
load_f1 = 0.942951438000943
real_f1 = 0.9411210551106923
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:17, 244.96it/s]

```

2.1.4 Amazon-Google - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Structured/Amazon-Google \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Amazon-Google.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/Amazon-Google \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level O2: FP16 training with FP32 batchnorm and FP32 master weights.

```

Defaults for this optimization level are:
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8496168255805969
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.06672077625989914
step: 20, loss: 0.37545186281204224
step: 30, loss: 0.36858952045440674
step: 40, loss: 0.4873368740081787
step: 50, loss: 0.2770627737045288
step: 60, loss: 0.3484615087509155
step: 70, loss: 0.20873525738716125
step: 80, loss: 0.3165645897388458
step: 90, loss: 0.36608582735061646
step: 100, loss: 0.1074187159538269
step: 110, loss: 0.363045871257782
step: 120, loss: 0.2290380299091339

```



```

step: 130, loss: 0.19129477441310883
step: 140, loss: 0.19874052703380585
step: 150, loss: 0.36176496744155884
step: 160, loss: 0.1904246062040329
step: 170, loss: 0.13125397264957428
step: 180, loss: 0.1597154438495636
step: 190, loss: 0.22314225137233734
step: 200, loss: 0.14433999359607697
step: 210, loss: 0.3474462628364563
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.6177606177606179, f1=0.6259842519685039,
best_f1=0.6259842519685039
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06143181025981903
step: 10, loss: 0.027963481843471527
step: 20, loss: 0.13224762678146362
step: 30, loss: 0.08556323498487473
step: 40, loss: 0.08185676485300064
step: 50, loss: 0.1992795169353485
step: 60, loss: 0.03967028111219406
step: 70, loss: 0.1315096914768219
step: 80, loss: 0.11534477770328522
step: 90, loss: 0.16667918860912323
step: 100, loss: 0.041226569563150406
step: 110, loss: 0.1529158055782318
step: 120, loss: 0.17942452430725098
step: 130, loss: 0.2050287127494812
step: 140, loss: 0.15301409363746643
step: 150, loss: 0.18412725627422333
step: 160, loss: 0.13133352994918823
step: 170, loss: 0.19719812273979187
step: 180, loss: 0.135127991437912
step: 190, loss: 0.085630401968956
step: 200, loss: 0.2650272846221924
step: 210, loss: 0.13032303750514984
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or

```

```

loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.6628787878787878, f1=0.6901960784313725,
best_f1=0.6901960784313725
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11370882391929626
step: 10, loss: 0.13864953815937042
step: 20, loss: 0.1695316880941391
step: 30, loss: 0.06474964320659637
step: 40, loss: 0.250325471162796
step: 50, loss: 0.04124877229332924
step: 60, loss: 0.3050420880317688
step: 70, loss: 0.073862724006176
step: 80, loss: 0.05188838019967079
step: 90, loss: 0.09945479780435562
step: 100, loss: 0.05433281883597374
step: 110, loss: 0.1122453510761261
step: 120, loss: 0.1911233514547348
step: 130, loss: 0.05964016914367676
step: 140, loss: 0.21296295523643494
step: 150, loss: 0.0743803158402443
step: 160, loss: 0.03961995244026184
step: 170, loss: 0.12833252549171448
step: 180, loss: 0.04298248514533043
step: 190, loss: 0.12683218717575073
step: 200, loss: 0.11582545191049576
step: 210, loss: 0.21689827740192413
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.6767485822306237, f1=0.6890130353817504,
best_f1=0.6890130353817504
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04547679424285889
step: 10, loss: 0.006543781608343124
step: 20, loss: 0.11726052314043045
step: 30, loss: 0.10262133926153183
step: 40, loss: 0.029261097311973572
step: 50, loss: 0.05150595307350159
step: 60, loss: 0.03259087726473808

```

```

step: 70, loss: 0.10445737093687057
step: 80, loss: 0.13706490397453308
step: 90, loss: 0.08278840780258179
step: 100, loss: 0.05843029171228409
step: 110, loss: 0.10345348715782166
step: 120, loss: 0.11824893951416016
step: 130, loss: 0.06623736768960953
step: 140, loss: 0.10750789940357208
step: 150, loss: 0.08336080610752106
step: 160, loss: 0.027424121275544167
step: 170, loss: 0.07789158821105957
step: 180, loss: 0.31039583683013916
step: 190, loss: 0.049324605613946915
step: 200, loss: 0.038560882210731506
step: 210, loss: 0.02335485816001892
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.6727605118829981, f1=0.6863468634686347,
best_f1=0.6890130353817504
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.031659629195928574
step: 10, loss: 0.044371165335178375
step: 20, loss: 0.011804888024926186
step: 30, loss: 0.008222163654863834
step: 40, loss: 0.11199745535850525
step: 50, loss: 0.09695973992347717
step: 60, loss: 0.01951838843524456
step: 70, loss: 0.12322448939085007
step: 80, loss: 0.06522577255964279
step: 90, loss: 0.03871653974056244
step: 100, loss: 0.043526820838451385
step: 110, loss: 0.01244918443262577
step: 120, loss: 0.024068240076303482
step: 130, loss: 0.022988859564065933
step: 140, loss: 0.02252386137843132
step: 150, loss: 0.0534214973449707
step: 160, loss: 0.029651494696736336
step: 170, loss: 0.010570341721177101
step: 180, loss: 0.10289215296506882
step: 190, loss: 0.0706983134150505
step: 200, loss: 0.07813755422830582
step: 210, loss: 0.014077706262469292
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.6792452830188679, f1=0.6876310272536688,
best_f1=0.6876310272536688
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09263630211353302
step: 10, loss: 0.030140377581119537
step: 20, loss: 0.015342140570282936
step: 30, loss: 0.0928516834974289
step: 40, loss: 0.004300189670175314
step: 50, loss: 0.004838981200009584
step: 60, loss: 0.0898156389594078
step: 70, loss: 0.0006408413173630834
step: 80, loss: 0.04881294444203377
step: 90, loss: 0.06659151613712311
step: 100, loss: 0.07983099669218063
step: 110, loss: 0.026197202503681183
step: 120, loss: 0.00763722974807024
step: 130, loss: 0.021987339481711388
step: 140, loss: 0.019171779975295067
step: 150, loss: 0.02712797187268734
step: 160, loss: 0.13037163019180298
step: 170, loss: 0.21230515837669373
step: 180, loss: 0.0008771262946538627
step: 190, loss: 0.05544690787792206
step: 200, loss: 0.012983138673007488
step: 210, loss: 0.008207147940993309
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.6759443339960238, f1=0.6987951807228915,
best_f1=0.6876310272536688
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003939175978302956
step: 10, loss: 0.061604734510183334
step: 20, loss: 0.03686133399605751
step: 30, loss: 0.012700216844677925
step: 40, loss: 0.05265402793884277
step: 50, loss: 0.022680504247546196
step: 60, loss: 0.0866294875741005
step: 70, loss: 0.030629508197307587
step: 80, loss: 0.004996820818632841

```

```

step: 90, loss: 0.010087395086884499
step: 100, loss: 0.031791072338819504
step: 110, loss: 0.014190465211868286
step: 120, loss: 0.02972988784313202
step: 130, loss: 0.003047942416742444
step: 140, loss: 0.0030615166760981083
step: 150, loss: 0.040007006376981735
step: 160, loss: 0.008808447048068047
step: 170, loss: 0.009883328340947628
step: 180, loss: 0.001616115216165781
step: 190, loss: 0.004431221168488264
step: 200, loss: 0.0009025362087413669
step: 210, loss: 0.021053949370980263
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.6831275720164609, f1=0.6804123711340206,
best_f1=0.6804123711340206
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008804019540548325
step: 10, loss: 0.0011139609850943089
step: 20, loss: 0.005345884244889021
step: 30, loss: 0.004380377475172281
step: 40, loss: 0.011359019204974174
step: 50, loss: 0.16211016476154327
step: 60, loss: 0.08686735481023788
step: 70, loss: 0.005033530294895172
step: 80, loss: 0.028460126370191574
step: 90, loss: 0.013232748955488205
step: 100, loss: 0.0007814109558239579
step: 110, loss: 0.007426361087709665
step: 120, loss: 0.0030052405782043934
step: 130, loss: 0.06202731281518936
step: 140, loss: 0.0008177506970241666
step: 150, loss: 0.05813148617744446
step: 160, loss: 0.0870639905333519
step: 170, loss: 0.02120787650346756
step: 180, loss: 0.061134591698646545
step: 190, loss: 0.05319582670927048
step: 200, loss: 0.04678528383374214
step: 210, loss: 0.05965966358780861
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.682261208576998, f1=0.682261208576998,
best_f1=0.6804123711340206
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010359928011894226
step: 10, loss: 0.019760098308324814
step: 20, loss: 0.006887355819344521
step: 30, loss: 0.025813238695263863
step: 40, loss: 0.06250999122858047
step: 50, loss: 0.0012366451555863023
step: 60, loss: 0.000735917710699141
step: 70, loss: 0.00121121178381145
step: 80, loss: 0.0008405921398662031
step: 90, loss: 0.05666746571660042
step: 100, loss: 0.0405084565281868
step: 110, loss: 0.000284229579847306
step: 120, loss: 0.002759899478405714
step: 130, loss: 0.0006268433062359691
step: 140, loss: 0.024184929206967354
step: 150, loss: 0.004708807449787855
step: 160, loss: 0.0027284862007945776
step: 170, loss: 0.0050551691092550755
step: 180, loss: 0.014987895265221596
step: 190, loss: 0.039881300181150436
step: 200, loss: 0.0008581542642787099
step: 210, loss: 0.059044547379016876
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.6785714285714285, f1=0.6944444444444444,
best_f1=0.6804123711340206
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005241211038082838
step: 10, loss: 0.0018778101075440645
step: 20, loss: 0.007121308706700802
step: 30, loss: 0.004853669088333845
step: 40, loss: 0.03150811046361923
step: 50, loss: 0.000496042484883219
step: 60, loss: 0.00027460954152047634
step: 70, loss: 0.0012074195547029376
step: 80, loss: 9.938109724316746e-05
step: 90, loss: 0.0047198208048939705
step: 100, loss: 0.0003836624091491103

```

```

step: 110, loss: 0.005939028691500425
step: 120, loss: 0.003250119974836707
step: 130, loss: 0.0008315385784953833
step: 140, loss: 0.0002716912713367492
step: 150, loss: 0.036770012229681015
step: 160, loss: 0.03013644739985466
step: 170, loss: 0.011678916402161121
step: 180, loss: 0.02556227333843708
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 190, loss: 0.15504345297813416
step: 200, loss: 0.028506074100732803
step: 210, loss: 0.004401540849357843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.6777163904235728, f1=0.6981132075471699,
best_f1=0.6804123711340206
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02129356563091278
step: 10, loss: 0.0034738206304609776
step: 20, loss: 0.0005949687329120934
step: 30, loss: 0.00138366880428791
step: 40, loss: 0.007191253826022148
step: 50, loss: 0.01699271984398365
step: 60, loss: 0.016089286655187607
step: 70, loss: 0.003933326341211796
step: 80, loss: 0.03125154599547386
step: 90, loss: 0.00033320928923785686
step: 100, loss: 0.0027961626183241606
step: 110, loss: 0.0012939546722918749
step: 120, loss: 0.00046048438525758684
step: 130, loss: 0.0900677740573883
step: 140, loss: 0.002121277851983905
step: 150, loss: 0.03718606382608414
step: 160, loss: 0.00038713589310646057
step: 170, loss: 0.08687881380319595
step: 180, loss: 0.01609773002564907
step: 190, loss: 0.00319555401802063
step: 200, loss: 0.00022659706883132458
step: 210, loss: 0.0004618769162334502
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.

```

```

    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 11: dev_f1=0.6846153846153845, f1=0.6904761904761906,
best_f1=0.6904761904761906
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002979272685479373
step: 10, loss: 0.00040750097832642496
step: 20, loss: 0.0004118247306905687
step: 30, loss: 0.00033880159026011825
step: 40, loss: 0.010880833491683006
step: 50, loss: 0.00937323085963726
step: 60, loss: 0.010474927723407745
step: 70, loss: 0.00020018433860968798
step: 80, loss: 0.00045802941895090044
step: 90, loss: 0.00010077776096295565
step: 100, loss: 0.0002720133343245834
step: 110, loss: 0.0009381513809785247
step: 120, loss: 0.0002955296076834202
step: 130, loss: 0.005241443403065205
step: 140, loss: 0.0015634509036317468
step: 150, loss: 0.00019608493312262
step: 160, loss: 0.0013872751733288169
step: 170, loss: 0.005040793213993311
step: 180, loss: 0.0005308658583089709
step: 190, loss: 0.06319388002157211
step: 200, loss: 0.09222712367773056
step: 210, loss: 0.0009113645064644516
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.6666666666666666, f1=0.693446088794926,
best_f1=0.6904761904761906
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00022882389021106064
step: 10, loss: 0.0033274057786911726
step: 20, loss: 0.004695525858551264
step: 30, loss: 0.011869711801409721
step: 40, loss: 0.0004898713086731732
step: 50, loss: 0.00016300873539876193
step: 60, loss: 0.005571757908910513
step: 70, loss: 0.020321233198046684
step: 80, loss: 0.03134618699550629
step: 90, loss: 0.05453154072165489
step: 100, loss: 0.02941843681037426
step: 110, loss: 0.004025366622954607

```



```

step: 120, loss: 0.006669190712273121
step: 130, loss: 0.0003290932800155133
step: 140, loss: 0.0020930198952555656
step: 150, loss: 0.00012299380614422262
step: 160, loss: 0.00044785827049054205
step: 170, loss: 0.00029025872936472297
step: 180, loss: 0.006778988521546125
step: 190, loss: 0.00022279545373748988
step: 200, loss: 0.00015731569146737456
step: 210, loss: 0.00022253612405620515
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.6752688172043011, f1=0.7038626609442059,
best_f1=0.6904761904761906
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09498266130685806
step: 10, loss: 0.0001486620312789455
step: 20, loss: 0.0001543869002489373
step: 30, loss: 0.004450892563909292
step: 40, loss: 0.014974667690694332
step: 50, loss: 0.0006112222908996046
step: 60, loss: 0.005244655534625053
step: 70, loss: 0.003749546827748418
step: 80, loss: 0.025214003399014473
step: 90, loss: 0.0003033960238099098
step: 100, loss: 0.00024490992655046284
step: 110, loss: 0.04767480865120888
step: 120, loss: 0.012989719398319721
step: 130, loss: 0.0001539344375487417
step: 140, loss: 0.00010696947720134631
step: 150, loss: 0.00027080296422354877
step: 160, loss: 0.00028824570472352207
step: 170, loss: 0.0004615538928192109
step: 180, loss: 0.0005021885153837502
step: 190, loss: 0.00012165845691924915
step: 200, loss: 0.00010851520346477628
step: 210, loss: 0.0009143349016085267
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.6711409395973156, f1=0.6784922394678493,
best_f1=0.6904761904761906
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.00446130707859993
step: 10, loss: 0.01292375847697258
step: 20, loss: 0.0001142206383519806
step: 30, loss: 0.00018457254918757826
step: 40, loss: 0.00012041867739753798
step: 50, loss: 0.00018221695790998638
step: 60, loss: 0.006975241471081972
step: 70, loss: 0.00014024043048266321
step: 80, loss: 0.00043342309072613716
step: 90, loss: 0.0002793285239022225
step: 100, loss: 0.0006154453731141984
step: 110, loss: 8.693914423929527e-05
step: 120, loss: 0.00026320599135942757
step: 130, loss: 0.004131821449846029
step: 140, loss: 0.0011817073682323098
step: 150, loss: 0.00014019591617397964
step: 160, loss: 0.008513827808201313
step: 170, loss: 0.054172564297914505
step: 180, loss: 0.0006976725999265909
step: 190, loss: 0.0001286643964704126
step: 200, loss: 0.00019170729501638561
step: 210, loss: 0.0027904666494578123
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.6776180698151951, f1=0.7012448132780082,
best_f1=0.6904761904761906
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic

```

```
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2293it [00:06, 340.43it/s]
load_f1 = 0.684931506849315
real_f1 = 0.6820809248554913
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:18, 241.49it/s]
```

2.1.5 Walmart-Amazon - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
    --task Structured/Walmart-Amazon \
    --batch_size 32 \
    --max_len 256 \
    --lr 3e-5 \
    --n_epochs 15 \
    --finetuning \
    --lm distilbert \
    --fp16 \
    --save_model

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Walmart-Amazon.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
    --task Structured/Walmart-Amazon \
    --input_path input/input_small.jsonl \
    --output_path output/output_small.jsonl \
    --lm distilbert \
    --use_gpu \
    --fp16 \
    --checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:

UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`. In PyTorch 1.1.0 and later, you should call them in the opposite order: `optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

```
"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
```

```
step: 0, loss: 0.8669355511665344
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0

```
step: 10, loss: 0.17188656330108643
```

```
step: 20, loss: 0.15904809534549713
```

```
step: 30, loss: 0.509552001953125
```

```
step: 40, loss: 0.2554256319999695
```

```
step: 50, loss: 0.3112546503543854
```

```
step: 60, loss: 0.3579767346382141
```

```
step: 70, loss: 0.1766635924577713
```

```
step: 80, loss: 0.5001188516616821
```

```
step: 90, loss: 0.22812311351299286
```

```
step: 100, loss: 0.22025740146636963
```

```

step: 110, loss: 0.23477089405059814
step: 120, loss: 0.4248694181442261
step: 130, loss: 0.3442996144294739
step: 140, loss: 0.3377842307090759
step: 150, loss: 0.2677720785140991
step: 160, loss: 0.22442469000816345
step: 170, loss: 0.3576165735721588
step: 180, loss: 0.31458839774131775
step: 190, loss: 0.11499307304620743
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.4175257731958763, f1=0.3958333333333333,
best_f1=0.3958333333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.29960450530052185
step: 10, loss: 0.033939991146326065
step: 20, loss: 0.11375188827514648
step: 30, loss: 0.16025231778621674
step: 40, loss: 0.4475383162498474
step: 50, loss: 0.29738929867744446
step: 60, loss: 0.2129300832748413
step: 70, loss: 0.24103158712387085
step: 80, loss: 0.17825254797935486
step: 90, loss: 0.13535283505916595
step: 100, loss: 0.2911425530910492
step: 110, loss: 0.15309534966945648
step: 120, loss: 0.2733123004436493
step: 130, loss: 0.13878236711025238
step: 140, loss: 0.17838628590106964
step: 150, loss: 0.02144622430205345
step: 160, loss: 0.02709062583744526
step: 170, loss: 0.240775465965271
step: 180, loss: 0.24531243741512299
step: 190, loss: 0.12697678804397583
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)

```

```

epoch 2: dev_f1=0.7293447293447294, f1=0.7514124293785311,
best_f1=0.7514124293785311
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.19187025725841522
step: 10, loss: 0.31436145305633545
step: 20, loss: 0.02945888228714466
step: 30, loss: 0.055899862200021744
step: 40, loss: 0.023944688960909843
step: 50, loss: 0.1954934000968933
step: 60, loss: 0.02410297654569149
step: 70, loss: 0.1119682788848877
step: 80, loss: 0.2366180568933487
step: 90, loss: 0.04782779514789581
step: 100, loss: 0.10068083554506302
step: 110, loss: 0.18620085716247559
step: 120, loss: 0.016221528872847557
step: 130, loss: 0.05990636348724365
step: 140, loss: 0.05407818406820297
step: 150, loss: 0.11933362483978271
step: 160, loss: 0.09328490495681763
step: 170, loss: 0.05730729550123215
step: 180, loss: 0.02022542431950569
step: 190, loss: 0.1380031555891037
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.768, f1=0.7675070028011204, best_f1=0.7675070028011204
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02573634497821331
step: 10, loss: 0.06000306457281113
step: 20, loss: 0.017846576869487762
step: 30, loss: 0.02295149862766266
step: 40, loss: 0.004239444620907307
step: 50, loss: 0.11394940316677094
step: 60, loss: 0.08304721862077713
step: 70, loss: 0.013625379651784897
step: 80, loss: 0.012139901518821716
step: 90, loss: 0.00965818576514721
step: 100, loss: 0.02489384263753891
step: 110, loss: 0.024441536515951157

```

```

step: 120, loss: 0.02732408046722412
step: 130, loss: 0.23409348726272583
step: 140, loss: 0.03332964703440666
step: 150, loss: 0.039571259170770645
step: 160, loss: 0.02738470584154129
step: 170, loss: 0.028564972802996635
step: 180, loss: 0.20649248361587524
step: 190, loss: 0.08732262253761292
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.7692307692307693, f1=0.7810650887573964,
best_f1=0.7810650887573964
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0440429225564003
step: 10, loss: 0.00587152224034071
step: 20, loss: 0.04160308837890625
step: 30, loss: 0.01114022359251976
step: 40, loss: 0.017799746245145798
step: 50, loss: 0.008104735985398293
step: 60, loss: 0.02813524194061756
step: 70, loss: 0.04680860415101051
step: 80, loss: 0.009652023203670979
step: 90, loss: 0.10581475496292114
step: 100, loss: 0.052418339997529984
step: 110, loss: 0.010020031593739986
step: 120, loss: 0.0014185721520334482
step: 130, loss: 0.006472067907452583
step: 140, loss: 0.005426058545708656
step: 150, loss: 0.024074045941233635
step: 160, loss: 0.0023501880932599306
step: 170, loss: 0.007697932422161102
step: 180, loss: 0.06789000332355499
step: 190, loss: 0.047817476093769073
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.7905759162303665, f1=0.8053333333333333,

```

```

best_f1=0.8053333333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10401783883571625
step: 10, loss: 0.0006764632998965681
step: 20, loss: 0.02719777449965477
step: 30, loss: 0.007722924463450909
step: 40, loss: 0.021511273458600044
step: 50, loss: 0.008121688850224018
step: 60, loss: 0.07293180376291275
step: 70, loss: 0.04094063118100166
step: 80, loss: 0.05815044417977333
step: 90, loss: 0.1121237725019455
step: 100, loss: 0.11368156224489212
step: 110, loss: 0.14082561433315277
step: 120, loss: 0.02148781344294548
step: 130, loss: 0.0019099004566669464
step: 140, loss: 0.004207505378872156
step: 150, loss: 0.09693381935358047
step: 160, loss: 0.009518472477793694
step: 170, loss: 0.08806511014699936
step: 180, loss: 0.0065173134207725525
step: 190, loss: 0.014169500209391117
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.7754010695187165, f1=0.7780821917808218,
best_f1=0.8053333333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001738185528665781
step: 10, loss: 0.05124560371041298
step: 20, loss: 0.000946300511714071
step: 30, loss: 0.0036701723001897335
step: 40, loss: 0.0060212453827261925
step: 50, loss: 0.022392043843865395
step: 60, loss: 0.008839013054966927
step: 70, loss: 0.003977970685809851
step: 80, loss: 0.007791189011186361
step: 90, loss: 0.027766266837716103
step: 100, loss: 0.0012075708946213126
step: 110, loss: 0.015446977689862251
step: 120, loss: 0.0037881687749177217
step: 130, loss: 0.013746604323387146
step: 140, loss: 0.0012059389846399426
step: 150, loss: 0.002270694822072983

```



```

step: 160, loss: 0.0020309581886976957
step: 170, loss: 0.0018044551834464073
step: 180, loss: 0.01757226511836052
step: 190, loss: 0.0016624851850792766
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.7772020725388601, f1=0.7924528301886792,
best_f1=0.8053333333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006276262574829161
step: 10, loss: 0.0009844093583524227
step: 20, loss: 0.0016602990217506886
step: 30, loss: 0.004180642310529947
step: 40, loss: 0.0030561406165361404
step: 50, loss: 0.0004688969347625971
step: 60, loss: 0.0004879781627096236
step: 70, loss: 0.0031399631407111883
step: 80, loss: 0.020203880965709686
step: 90, loss: 0.0005556109827011824
step: 100, loss: 0.018614748492836952
step: 110, loss: 0.00046436130651272833
step: 120, loss: 0.0002981493016704917
step: 130, loss: 0.0015501162270084023
step: 140, loss: 0.002018808154389262
step: 150, loss: 0.0004965874250046909
step: 160, loss: 0.0011426376877352595
step: 170, loss: 0.0007868416141718626
step: 180, loss: 0.0006512743420898914
step: 190, loss: 0.019246160984039307
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.7804878048780487, f1=0.8021978021978022,
best_f1=0.8053333333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006478248978964984
step: 10, loss: 0.07159602642059326
step: 20, loss: 0.0033114482648670673
step: 30, loss: 0.022553445771336555
step: 40, loss: 0.0050486112013459206
step: 50, loss: 0.004254107363522053
step: 60, loss: 0.0013804933987557888
step: 70, loss: 0.0006612192955799401

```

```

step: 80, loss: 0.000679034914355725
step: 90, loss: 0.006845113355666399
step: 100, loss: 0.0010424783686175942
step: 110, loss: 0.0007254167576320469
step: 120, loss: 0.0186068806797266
step: 130, loss: 0.004322111140936613
step: 140, loss: 0.014178009703755379
step: 150, loss: 0.012944458052515984
step: 160, loss: 0.0008939707768149674
step: 170, loss: 0.002526169875636697
step: 180, loss: 0.004296178463846445
step: 190, loss: 0.0012761151883751154
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.7687861271676301, f1=0.7715133531157269,
best_f1=0.8053333333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00023872414021752775
step: 10, loss: 0.0005410074372775853
step: 20, loss: 0.0011061482364311814
step: 30, loss: 0.000804990588221699
step: 40, loss: 0.0023805273231118917
step: 50, loss: 0.00019384436018299311
step: 60, loss: 0.011961985379457474
step: 70, loss: 0.013619300909340382
step: 80, loss: 0.00021110598754603416
step: 90, loss: 0.007243643514811993
step: 100, loss: 0.001359514077194035
step: 110, loss: 0.03319401666522026
step: 120, loss: 0.020362555980682373
step: 130, loss: 0.0008795431931503117
step: 140, loss: 0.000547960342373699
step: 150, loss: 0.00018448711489327252
step: 160, loss: 0.0002500773698557168
step: 170, loss: 0.0005120935966260731
step: 180, loss: 0.00030641554621979594
step: 190, loss: 0.0006443533347919583
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.7726027397260273, f1=0.7808988764044944,
best_f1=0.8053333333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.00023025949485599995
step: 10, loss: 0.00036087672924622893
step: 20, loss: 0.0008920415421016514
step: 30, loss: 0.013060844503343105
step: 40, loss: 0.0002190673549193889
step: 50, loss: 0.0003734001948032528
step: 60, loss: 0.002090894151479006
step: 70, loss: 0.002791157690808177
step: 80, loss: 0.0002213277475675568
step: 90, loss: 0.0038763389457017183
step: 100, loss: 0.0008028086158446968
step: 110, loss: 0.00012668738781940192
step: 120, loss: 0.0009575230069458485
step: 130, loss: 0.0003953904961235821
step: 140, loss: 0.00014049957098904997
step: 150, loss: 0.0001901882205856964
step: 160, loss: 0.0001940647780429572
step: 170, loss: 0.0014448012225329876
step: 180, loss: 0.053243476897478104
step: 190, loss: 0.001079268753528595
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.7461139896373058, f1=0.7795698924731183,
best_f1=0.8053333333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0033867880702018738
step: 10, loss: 0.0014693960547447205
step: 20, loss: 0.0008284367504529655
step: 30, loss: 0.003876968054100871
step: 40, loss: 0.0002956808893941343
step: 50, loss: 0.1145927682518959
step: 60, loss: 0.00021288328571245074
step: 70, loss: 0.0007069525308907032
step: 80, loss: 0.000387580570532009
step: 90, loss: 0.002279933076351881
step: 100, loss: 0.0006663820240646601
step: 110, loss: 0.00033508785418234766
step: 120, loss: 0.0002870288444682956
step: 130, loss: 0.00010988224676111713
step: 140, loss: 0.2186799794435501
step: 150, loss: 0.0002271068369736895
step: 160, loss: 0.013343733735382557
step: 170, loss: 0.00036587691283784807
step: 180, loss: 0.0007429832476191223
step: 190, loss: 0.05095687508583069

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.752, f1=0.7692307692307692, best_f1=0.8053333333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00031356560066342354
step: 10, loss: 0.0027590885292738676
step: 20, loss: 0.0004668066103477031
step: 30, loss: 0.00021651627321261913
step: 40, loss: 0.00039928534533828497
step: 50, loss: 0.000548149342648685
step: 60, loss: 0.00014260852185543627
step: 70, loss: 0.00041516803321428597
step: 80, loss: 0.0017567496979609132
step: 90, loss: 0.00043442921014502645
step: 100, loss: 0.0004901381325908005
step: 110, loss: 0.000337209552526474
step: 120, loss: 0.0010911609278991818
step: 130, loss: 0.0012156963348388672
step: 140, loss: 0.00023675711418036371
step: 150, loss: 0.0001949869911186397
step: 160, loss: 0.006282950285822153
step: 170, loss: 0.0007850961992517114
step: 180, loss: 0.001002808567136526
step: 190, loss: 0.0027900540735572577
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.7586206896551724, f1=0.7828418230563003,
best_f1=0.8053333333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007139587542042136
step: 10, loss: 0.0003623895754572004
step: 20, loss: 0.012448370456695557
step: 30, loss: 0.0009593200520612299
step: 40, loss: 0.00014724236098118126
step: 50, loss: 0.000448095437604934
step: 60, loss: 0.0003674957843031734
step: 70, loss: 0.0004532648890744895
step: 80, loss: 0.00010019849287346005
step: 90, loss: 0.0003350837796460837
step: 100, loss: 0.00014082273992244154
step: 110, loss: 0.0003700325032696128
step: 120, loss: 0.0001738642022246495

```

```

step: 130, loss: 8.230486128013581e-05
step: 140, loss: 0.00042634078999981284
step: 150, loss: 0.00016557851631660014
step: 160, loss: 0.00024477409897372127
step: 170, loss: 0.055274009704589844
step: 180, loss: 0.00028892207774333656
step: 190, loss: 0.00291195185855031
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.7613941018766757, f1=0.7923497267759563,
best_f1=0.8053333333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007104522082954645
step: 10, loss: 0.00020054163178429008
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 20, loss: 0.043704211711883545
step: 30, loss: 0.00038262101588770747
step: 40, loss: 0.00015370426990557462
step: 50, loss: 0.00024659981136210263
step: 60, loss: 0.0003456637787166983
step: 70, loss: 0.0014764778316020966
step: 80, loss: 0.00047858021571300924
step: 90, loss: 0.0005172890378162265
step: 100, loss: 0.00019351149967405945
step: 110, loss: 0.0004401286714710295
step: 120, loss: 0.003703756956383586
step: 130, loss: 0.00020381489594001323
step: 140, loss: 0.009845229797065258
step: 150, loss: 0.00018534762784838676
step: 160, loss: 0.0003382846771273762
step: 170, loss: 0.0021568117663264275
step: 180, loss: 0.00039824831765145063
step: 190, loss: 0.008892803452908993
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.7631578947368421, f1=0.7903225806451613,
best_f1=0.8053333333333333
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

```

```

Defaults for this optimization level are:
enabled          : True
opt_level        : 02

```

```

cast_model_type      : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32  : True
master_weights       : True
loss_scale            : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled              : True
opt_level            : 02
cast_model_type      : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32  : True
master_weights       : True
loss_scale            : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
2049it [00:09, 223.00it/s]
load_f1 = 0.8010471204188483
real_f1 = 0.7819148936170213
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:17, 246.63it/s]

```

2.2 BASELINE TEXTUAL

2.2.1 Abt-Buy - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Textual/Abt-Buy \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \

```

```

--lm distilbert \
--fp16 \
--save_model

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Abt-Buy.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
--task Textual/Abt-Buy \
--input_path input/input_small.jsonl \
--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:

UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:

`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

"<https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>",
UserWarning)

step: 0, loss: 0.847989022731781

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0

step: 10, loss: 0.22609898447990417

step: 20, loss: 0.1614876389503479

step: 30, loss: 0.23020446300506592

step: 40, loss: 0.30970045924186707

step: 50, loss: 0.3737640976905823

step: 60, loss: 0.4373324513435364

step: 70, loss: 0.2980610430240631

step: 80, loss: 0.26018738746643066

step: 90, loss: 0.3855542540550232

step: 100, loss: 0.22431883215904236

step: 110, loss: 0.18076828122138977

step: 120, loss: 0.5781480669975281

step: 130, loss: 0.39658209681510925

step: 140, loss: 0.4782811999320984

step: 150, loss: 0.12809813022613525

step: 160, loss: 0.3666628301143646

step: 170, loss: 0.22462303936481476

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

warnings.warn("An input tensor was not cuda.")

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:

UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.

warnings.warn(SAVE_STATE_WARNING, UserWarning)

epoch 1: dev_f1=0.45049504950495056, f1=0.40201005025125625,

best_f1=0.40201005025125625

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

warnings.warn("An input tensor was not cuda.")

step: 0, loss: 0.34516313672065735

step: 10, loss: 0.12002266198396683

step: 20, loss: 0.31806281208992004

step: 30, loss: 0.15248075127601624

step: 40, loss: 0.18088079988956451

step: 50, loss: 0.28184783458709717

step: 60, loss: 0.23238034546375275

step: 70, loss: 0.1635378748178482

step: 80, loss: 0.15589137375354767

step: 90, loss: 0.14870685338974

step: 100, loss: 0.21391715109348297


```

step: 110, loss: 0.1799173802137375
step: 120, loss: 0.07873253524303436
step: 130, loss: 0.06707936525344849
step: 140, loss: 0.1262737661600113
step: 150, loss: 0.044730205088853836
step: 160, loss: 0.15251760184764862
step: 170, loss: 0.04560223966836929
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.7632183908045976, f1=0.7460674157303371,
best_f1=0.7460674157303371
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0494636595249176
step: 10, loss: 0.04467487707734108
step: 20, loss: 0.06249421462416649
step: 30, loss: 0.057450708001852036
step: 40, loss: 0.013742412440478802
step: 50, loss: 0.19404257833957672
step: 60, loss: 0.15721248090267181
step: 70, loss: 0.04879706725478172
step: 80, loss: 0.13141049444675446
step: 90, loss: 0.10150701552629471
step: 100, loss: 0.02525487169623375
step: 110, loss: 0.014854182489216328
step: 120, loss: 0.0031291688792407513
step: 130, loss: 0.12331197410821915
step: 140, loss: 0.005608092062175274
step: 150, loss: 0.055287666618824005
step: 160, loss: 0.1612185686826706
step: 170, loss: 0.2007627785205841
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.7950617283950617, f1=0.7817745803357313,
best_f1=0.7817745803357313
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08499693125486374
step: 10, loss: 0.12991315126419067
step: 20, loss: 0.016261056065559387
step: 30, loss: 0.02757357992231846
step: 40, loss: 0.00843943003565073
step: 50, loss: 0.0072571407072246075
step: 60, loss: 0.1495947539806366
step: 70, loss: 0.01927308551967144
step: 80, loss: 0.017332719638943672
step: 90, loss: 0.04116520658135414
step: 100, loss: 0.04963294416666031
step: 110, loss: 0.04667648673057556
step: 120, loss: 0.10326740145683289
step: 130, loss: 0.11121868342161179
step: 140, loss: 0.009956520982086658
step: 150, loss: 0.007880723103880882
step: 160, loss: 0.07429856061935425
step: 170, loss: 0.036149006336927414
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.7883211678832117, f1=0.7793427230046948,
best_f1=0.7817745803357313
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.013415823690593243
step: 10, loss: 0.02992870658636093
step: 20, loss: 0.017218783497810364
step: 30, loss: 0.03938927873969078
step: 40, loss: 0.08158403635025024
step: 50, loss: 0.0031694862991571426
step: 60, loss: 0.002849261276423931
step: 70, loss: 0.005021706223487854
step: 80, loss: 0.010474206879734993
step: 90, loss: 0.010082053020596504
step: 100, loss: 0.01681330054998398
step: 110, loss: 0.02026713453233242
step: 120, loss: 0.050721365958452225
step: 130, loss: 0.007735064718872309
step: 140, loss: 0.012960332445800304
step: 150, loss: 0.006385082378983498
step: 160, loss: 0.02750515006482601
step: 170, loss: 0.03483208268880844
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.7990654205607476, f1=0.779816513761468,
best_f1=0.779816513761468
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008107389323413372
step: 10, loss: 0.07252618670463562
step: 20, loss: 0.0036344430409371853
step: 30, loss: 0.0512414425611496
step: 40, loss: 0.006315434817224741
step: 50, loss: 0.010198112577199936
step: 60, loss: 0.055023208260536194
step: 70, loss: 0.012799343094229698
step: 80, loss: 0.022933926433324814
step: 90, loss: 0.021686969324946404
step: 100, loss: 0.16053171455860138
step: 110, loss: 0.006686143111437559
step: 120, loss: 0.09495677053928375
step: 130, loss: 0.07187750190496445
step: 140, loss: 0.005513145588338375
step: 150, loss: 0.18334701657295227
step: 160, loss: 0.06064321845769882
step: 170, loss: 0.00530611677095294
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.7951219512195121, f1=0.7732696897374701,
best_f1=0.779816513761468
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006808798760175705
step: 10, loss: 0.0013843659544363618
step: 20, loss: 0.005590041168034077
step: 30, loss: 0.004533444531261921
step: 40, loss: 0.006555183324962854
step: 50, loss: 0.0022354600951075554
step: 60, loss: 0.14361250400543213
step: 70, loss: 0.011502892710268497
step: 80, loss: 0.002047083107754588
step: 90, loss: 0.012900969944894314
step: 100, loss: 0.06256672739982605
step: 110, loss: 0.0011195713886991143
step: 120, loss: 0.1155313029885292

```

```

step: 130, loss: 0.020858291536569595
step: 140, loss: 0.030964244157075882
step: 150, loss: 0.04635734111070633
step: 160, loss: 0.006629447918385267
step: 170, loss: 0.16375236213207245
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.7868131868131867, f1=0.7543103448275862,
best_f1=0.779816513761468
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.021820614114403725
step: 10, loss: 0.006504920776933432
step: 20, loss: 0.033288221806287766
step: 30, loss: 0.06060699000954628
step: 40, loss: 0.0018203203799203038
step: 50, loss: 0.005024250131100416
step: 60, loss: 0.011104482226073742
step: 70, loss: 0.04216810688376427
step: 80, loss: 0.0018265163525938988
step: 90, loss: 0.10624925792217255
step: 100, loss: 0.030379116535186768
step: 110, loss: 0.03012370876967907
step: 120, loss: 0.0010323296301066875
step: 130, loss: 0.0005362665979191661
step: 140, loss: 0.0024221616331487894
step: 150, loss: 0.02134137600660324
step: 160, loss: 0.05765381455421448
step: 170, loss: 0.0060839103534817696
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.7931034482758621, f1=0.79136690647482,
best_f1=0.779816513761468
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0024647654499858618
step: 10, loss: 0.18013276159763336
step: 20, loss: 0.0007020572666078806
step: 30, loss: 0.046186164021492004
step: 40, loss: 0.01419773604720831
step: 50, loss: 0.0024186482187360525
step: 60, loss: 0.0005032534245401621
step: 70, loss: 0.007142565678805113
step: 80, loss: 0.035290226340293884

```

```

step: 90, loss: 0.010626868344843388
step: 100, loss: 0.017446039244532585
step: 110, loss: 0.00011716664448613301
step: 120, loss: 0.01100088655948639
step: 130, loss: 0.002576923929154873
step: 140, loss: 0.0014911438338458538
step: 150, loss: 0.016248973086476326
step: 160, loss: 0.013221034780144691
step: 170, loss: 0.02440495789051056
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.7695961995249406, f1=0.7777777777777778,
best_f1=0.779816513761468
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006219567148946226
step: 10, loss: 0.021814649924635887
step: 20, loss: 0.0007994438637979329
step: 30, loss: 0.000704342033714056
step: 40, loss: 0.008538352325558662
step: 50, loss: 0.006431201938539743
step: 60, loss: 0.002851779805496335
step: 70, loss: 0.0002986736362800002
step: 80, loss: 0.005375591106712818
step: 90, loss: 0.03484475612640381
step: 100, loss: 0.0005912878550589085
step: 110, loss: 0.0016916290624067187
step: 120, loss: 0.014351492747664452
step: 130, loss: 0.004427798558026552
step: 140, loss: 0.0021628644317388535
step: 150, loss: 0.0002635857090353966
step: 160, loss: 0.00013352233509067446
step: 170, loss: 0.0008482547709718347
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.7576470588235293, f1=0.7692307692307693,
best_f1=0.779816513761468
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002649854577612132
step: 10, loss: 0.06061531975865364
step: 20, loss: 0.03577844798564911
step: 30, loss: 0.0007355657871812582
step: 40, loss: 0.029545975849032402

```

```

step: 50, loss: 0.011468295939266682
step: 60, loss: 0.010030548088252544
step: 70, loss: 0.00031680933898314834
step: 80, loss: 0.14544542133808136
step: 90, loss: 0.0036194869317114353
step: 100, loss: 0.0006188798579387367
step: 110, loss: 0.02031656540930271
step: 120, loss: 0.006057279650121927
step: 130, loss: 0.0009572989074513316
step: 140, loss: 0.01812742091715336
step: 150, loss: 0.0006387550965882838
step: 160, loss: 0.006856993306428194
step: 170, loss: 0.04652588441967964
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.7666666666666666, f1=0.772093023255814,
best_f1=0.779816513761468
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005875055678188801
step: 10, loss: 0.0018624525982886553
step: 20, loss: 0.04472855105996132
step: 30, loss: 0.0011641904711723328
step: 40, loss: 0.00019849513773806393
step: 50, loss: 0.0013511305442079902
step: 60, loss: 0.00030510162468999624
step: 70, loss: 0.009929098188877106
step: 80, loss: 0.02682502008974552
step: 90, loss: 0.000323875603498891
step: 100, loss: 0.083231620490551
step: 110, loss: 0.0002933960349764675
step: 120, loss: 0.0006520376191474497
step: 130, loss: 0.0006321752443909645
step: 140, loss: 0.0003223212552256882
step: 150, loss: 0.0016747090267017484
step: 160, loss: 0.013792975805699825
step: 170, loss: 0.003939252346754074
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.7674418604651163, f1=0.7899543378995434,
best_f1=0.779816513761468
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00017902949184644967

```

```

step: 10, loss: 0.04345029965043068
step: 20, loss: 0.00037581208744086325
step: 30, loss: 0.00021539456793107092
step: 40, loss: 0.00010727845801739022
step: 50, loss: 0.0036073452793061733
step: 60, loss: 0.0003783387946896255
step: 70, loss: 0.013953618705272675
step: 80, loss: 0.002699014265090227
step: 90, loss: 0.0003398003173060715
step: 100, loss: 0.00035828049294650555
step: 110, loss: 0.0019799675792455673
step: 120, loss: 0.00022415058629121631
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 130, loss: 0.008967564441263676
step: 140, loss: 0.0016560280928388238
step: 150, loss: 0.00024105572083499283
step: 160, loss: 0.00018705794354900718
step: 170, loss: 0.0007487000548280776
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.7741935483870968, f1=0.7883211678832117,
best_f1=0.779816513761468
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002647611836437136
step: 10, loss: 0.0002028021845035255
step: 20, loss: 0.00251835398375988
step: 30, loss: 0.010045839473605156
step: 40, loss: 0.0006567424279637635
step: 50, loss: 0.001607157988473773
step: 60, loss: 0.0075134397484362125
step: 70, loss: 0.008939862251281738
step: 80, loss: 0.004021371714770794
step: 90, loss: 0.00035387053503654897
step: 100, loss: 0.0006287411670200527
step: 110, loss: 0.001113886944949627
step: 120, loss: 0.0004668939218390733
step: 130, loss: 0.01843615435063839
step: 140, loss: 0.0006763438577763736
step: 150, loss: 0.0009345213766209781
step: 160, loss: 0.026683615520596504
step: 170, loss: 0.0010875883745029569
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.7830188679245284, f1=0.7877358490566038,

```

```

best_f1=0.779816513761468
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004824867646675557
step: 10, loss: 0.00263602496124804
step: 20, loss: 0.006889014504849911
step: 30, loss: 0.026889920234680176
step: 40, loss: 0.0001416349405189976
step: 50, loss: 0.002641246886923909
step: 60, loss: 9.818377293413505e-05
step: 70, loss: 0.0013203778071328998
step: 80, loss: 0.00015985051868483424
step: 90, loss: 0.0001337589928880334
step: 100, loss: 0.0006304295384325087
step: 110, loss: 0.0003189587441738695
step: 120, loss: 0.027993835508823395
step: 130, loss: 0.001605743425898254
step: 140, loss: 0.0028795532416552305
step: 150, loss: 0.03580065444111824
step: 160, loss: 0.00014576740795746446
step: 170, loss: 0.0030678759794682264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.781725888324873, f1=0.7931034482758621,
best_f1=0.779816513761468
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic

```



```
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
1916it [00:05, 329.88it/s]
load_f1 = 0.7982062780269058
real_f1 = 0.7824175824175823
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:17, 253.95it/s]
```

2.3 BASELINE DIRTY

2.3.1 DBLP-ACM - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
    --task Dirty/DBLP-ACM \
    --batch_size 32 \
    --max_len 256 \
    --lr 3e-5 \
    --n_epochs 15 \
    --finetuning \
    --lm distilbert \
    --fp16 \
    --save_model

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-ACM.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
    --task Dirty/DBLP-ACM \
    --input_path input/input_small.jsonl \
    --output_path output/output_small.jsonl \
    --lm distilbert \
    --use_gpu \
```

```
--fp16 \  
--checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled          : True  
opt_level        : 02  
cast_model_type  : torch.float16  
patch_torch_functions : False  
keep_batchnorm_fp32 : True  
master_weights   : True  
loss_scale       : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True  
opt_level        : 02  
cast_model_type  : torch.float16  
patch_torch_functions : False  
keep_batchnorm_fp32 : True  
master_weights   : True  
loss_scale       : dynamic
```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:

UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`. In PyTorch 1.1.0 and later, you should call them in the opposite order: `optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

```
"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",  
UserWarning)
```

```
step: 0, loss: 0.8175169825553894
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0

```
step: 10, loss: 0.47238922119140625
```

```
step: 20, loss: 0.5943049788475037
```

```
step: 30, loss: 0.43794965744018555
```

```
step: 40, loss: 0.2084253430366516
```

```
step: 50, loss: 0.09218667447566986
```

```
step: 60, loss: 0.05129352584481239
```

```

step: 70, loss: 0.18333759903907776
step: 80, loss: 0.2555314302444458
step: 90, loss: 0.03569202125072479
step: 100, loss: 0.050880830734968185
step: 110, loss: 0.10270559042692184
step: 120, loss: 0.05573539808392525
step: 130, loss: 0.008887084200978279
step: 140, loss: 0.024609742686152458
step: 150, loss: 0.1604956090450287
step: 160, loss: 0.2609061300754547
step: 170, loss: 0.01875298097729683
step: 180, loss: 0.058728426694869995
step: 190, loss: 0.03425084426999092
step: 200, loss: 0.0035043135285377502
step: 210, loss: 0.00744800316169858
step: 220, loss: 0.014566808007657528
step: 230, loss: 0.024879666045308113
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9522752497225305, f1=0.9663677130044843,
best_f1=0.9663677130044843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.020284222438931465
step: 10, loss: 0.0013324638130143285
step: 20, loss: 0.011925051920115948
step: 30, loss: 0.001953759463503957
step: 40, loss: 0.013882302679121494
step: 50, loss: 0.03907591477036476
step: 60, loss: 0.010070309042930603
step: 70, loss: 0.013914735056459904
step: 80, loss: 0.003650134662166238
step: 90, loss: 0.006789719220250845
step: 100, loss: 0.13058240711688995
step: 110, loss: 0.11863704770803452
step: 120, loss: 0.028129681944847107
step: 130, loss: 0.02907893992960453
step: 140, loss: 0.2348058670759201
step: 150, loss: 0.0182882659137249
step: 160, loss: 0.00949602946639061
step: 170, loss: 0.04707279056310654
step: 180, loss: 0.009854442439973354

```

```

step: 190, loss: 0.11459428817033768
step: 200, loss: 0.012723837047815323
step: 210, loss: 0.06513965874910355
step: 220, loss: 0.0036702484358102083
step: 230, loss: 0.0074135055765509605
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9575892857142857, f1=0.9643652561247216,
best_f1=0.9643652561247216
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04462062194943428
step: 10, loss: 0.010111089795827866
step: 20, loss: 0.0014150870265439153
step: 30, loss: 0.006767385173588991
step: 40, loss: 0.011787359602749348
step: 50, loss: 0.0037815298419445753
step: 60, loss: 0.0010609915480017662
step: 70, loss: 0.0011126263998448849
step: 80, loss: 0.0017979068215936422
step: 90, loss: 0.005586197599768639
step: 100, loss: 0.0009572763228788972
step: 110, loss: 0.006705989595502615
step: 120, loss: 0.018170788884162903
step: 130, loss: 0.0007683485746383667
step: 140, loss: 0.0029593093786388636
step: 150, loss: 0.0010763893369585276
step: 160, loss: 0.3124660551548004
step: 170, loss: 0.01895003952085972
step: 180, loss: 0.008515244349837303
step: 190, loss: 0.010845710523426533
step: 200, loss: 0.01859602890908718
step: 210, loss: 0.00456791277974844
step: 220, loss: 0.014418695122003555
step: 230, loss: 0.1503172516822815
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)

```

```

epoch 3: dev_f1=0.9589345172031077, f1=0.9643652561247216,
best_f1=0.9643652561247216
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010158509016036987
step: 10, loss: 0.007075135130435228
step: 20, loss: 0.001435084268450737
step: 30, loss: 0.0032848981209099293
step: 40, loss: 0.0048149703070521355
step: 50, loss: 0.0019078637706115842
step: 60, loss: 0.033956415951251984
step: 70, loss: 0.02149461954832077
step: 80, loss: 0.056947290897369385
step: 90, loss: 0.0026656771078705788
step: 100, loss: 0.017727181315422058
step: 110, loss: 0.025219758972525597
step: 120, loss: 0.011287430301308632
step: 130, loss: 0.018230272457003593
step: 140, loss: 0.0002940950798802078
step: 150, loss: 0.0010394523851573467
step: 160, loss: 0.0004488896229304373
step: 170, loss: 0.020264994353055954
step: 180, loss: 0.067962147295475
step: 190, loss: 0.0006288600852712989
step: 200, loss: 0.0015487815253436565
step: 210, loss: 0.000619594007730484
step: 220, loss: 0.11075081676244736
step: 230, loss: 0.008704650215804577
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.9643652561247216, f1=0.9688888888888889,
best_f1=0.9688888888888889
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0015555715654045343
step: 10, loss: 0.002049270085990429
step: 20, loss: 0.0006892636301927269
step: 30, loss: 0.0009357171365991235
step: 40, loss: 0.0007433024002239108
step: 50, loss: 0.0006672046729363501
step: 60, loss: 0.00043249837472103536

```

```

step: 70, loss: 0.00017959099204745144
step: 80, loss: 0.006790655665099621
step: 90, loss: 0.0037496008444577456
step: 100, loss: 0.0008662066538818181
step: 110, loss: 0.000833208963740617
step: 120, loss: 0.033437520265579224
step: 130, loss: 0.012622818350791931
step: 140, loss: 0.0016579045914113522
step: 150, loss: 0.0002553025842644274
step: 160, loss: 0.016123196110129356
step: 170, loss: 0.1732107698917389
step: 180, loss: 0.009434442967176437
step: 190, loss: 0.0012282904936000705
step: 200, loss: 0.004455693066120148
step: 210, loss: 0.0007600939716212451
step: 220, loss: 0.008663229644298553
step: 230, loss: 0.0008860617526806891
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.9665924276169264, f1=0.972129319955407,
best_f1=0.972129319955407
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005782740656286478
step: 10, loss: 0.0007468833355233073
step: 20, loss: 0.0025869745295494795
step: 30, loss: 0.00025497074238955975
step: 40, loss: 0.010694184340536594
step: 50, loss: 0.034432731568813324
step: 60, loss: 0.022652557119727135
step: 70, loss: 0.0011288155801594257
step: 80, loss: 0.0003384357551112771
step: 90, loss: 0.0003343881689943373
step: 100, loss: 0.00019059285114053637
step: 110, loss: 0.0342380553483963
step: 120, loss: 0.00046765845036134124
step: 130, loss: 0.00016723823500797153
step: 140, loss: 0.000394967821193859
step: 150, loss: 0.0006135173607617617
step: 160, loss: 0.004011299926787615
step: 170, loss: 6.896178092574701e-05
step: 180, loss: 0.00027279602363705635

```

```

step: 190, loss: 0.000973376736510545
step: 200, loss: 0.0001723304158076644
step: 210, loss: 0.005525696091353893
step: 220, loss: 0.00012541103933472186
step: 230, loss: 0.0002982829755637795
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.9656699889258028, f1=0.9711111111111111,
best_f1=0.972129319955407
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07429412752389908
step: 10, loss: 0.00010390251554781571
step: 20, loss: 0.0004894064622931182
step: 30, loss: 0.0005473305936902761
step: 40, loss: 0.003091041697189212
step: 50, loss: 0.00039235129952430725
step: 60, loss: 0.04409119859337807
step: 70, loss: 0.0007505387184210122
step: 80, loss: 0.00039782762178219855
step: 90, loss: 0.0005277722957544029
step: 100, loss: 0.00033075985265895724
step: 110, loss: 0.05569605529308319
step: 120, loss: 0.0003294929920230061
step: 130, loss: 0.004336063284426928
step: 140, loss: 0.0003522555634845048
step: 150, loss: 0.000534545979462564
step: 160, loss: 0.00017031685274560004
step: 170, loss: 0.00010858845053007826
step: 180, loss: 0.0002494274522177875
step: 190, loss: 0.19179315865039825
step: 200, loss: 0.002145901322364807
step: 210, loss: 0.00012954568956047297
step: 220, loss: 0.008606700226664543
step: 230, loss: 0.024233417585492134
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.9684684684684683, f1=0.9697648376259798,
best_f1=0.9697648376259798
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.020489495247602463
step: 10, loss: 0.050627727061510086
step: 20, loss: 0.0002834798360709101
step: 30, loss: 0.015525023452937603
step: 40, loss: 0.0008150136563926935
step: 50, loss: 0.0017703288467600942
step: 60, loss: 0.00013198213127907366
step: 70, loss: 0.00018102822650689632
step: 80, loss: 0.00047533237375319004
step: 90, loss: 0.0009628729894757271
step: 100, loss: 0.00012586273078341037
step: 110, loss: 0.00015911956143099815
step: 120, loss: 0.00010630799806676805
step: 130, loss: 0.0007759131258353591
step: 140, loss: 7.186598668340594e-05
step: 150, loss: 6.334621866699308e-05
step: 160, loss: 0.0007150995661504567
step: 170, loss: 0.085594043135643
step: 180, loss: 0.007840093225240707
step: 190, loss: 0.00025658775120973587
step: 200, loss: 0.002159304916858673
step: 210, loss: 0.0012570740655064583
step: 220, loss: 0.00012903647439088672
step: 230, loss: 0.0478690080344677
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.967452300785634, f1=0.967741935483871,
best_f1=0.9697648376259798
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003449928481131792
step: 10, loss: 0.000488129910081625
step: 20, loss: 0.0017869111616164446
step: 30, loss: 0.001624140772037208
step: 40, loss: 0.0036205914802849293
step: 50, loss: 0.0003838293778244406
step: 60, loss: 0.0033573440741747618
step: 70, loss: 0.001191129325889051
step: 80, loss: 0.03224599361419678
step: 90, loss: 0.0013081127544865012
step: 100, loss: 0.0007088822312653065
step: 110, loss: 0.0002195495180785656
step: 120, loss: 0.023623205721378326
step: 130, loss: 0.00020011138985864818
step: 140, loss: 0.010375584475696087

```



```

step: 150, loss: 6.291902536759153e-05
step: 160, loss: 0.0008517776150256395
step: 170, loss: 0.00018102272588294
step: 180, loss: 0.00019441210315562785
step: 190, loss: 0.0006436730618588626
step: 200, loss: 0.0015270261792466044
step: 210, loss: 0.00021694961469620466
step: 220, loss: 0.03209687024354935
step: 230, loss: 0.007338333874940872
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 9: dev_f1=0.968609865470852, f1=0.9707865168539327,
best_f1=0.9707865168539327
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00018596994050312787
step: 10, loss: 8.07570104370825e-05
step: 20, loss: 0.001677132211625576
step: 30, loss: 0.0004348820657469332
step: 40, loss: 0.00016323250019922853
step: 50, loss: 0.00035456495243124664
step: 60, loss: 0.062232717871665955
step: 70, loss: 0.0007058964110910892
step: 80, loss: 0.0002995004178956151
step: 90, loss: 0.0004034129960928112
step: 100, loss: 0.0002328384725842625
step: 110, loss: 0.00011168546188855544
step: 120, loss: 0.019341683015227318
step: 130, loss: 0.0005636637797579169
step: 140, loss: 0.011397193185985088
step: 150, loss: 0.0001825929939514026
step: 160, loss: 0.00012172883725725114
step: 170, loss: 0.00039911147905513644
step: 180, loss: 0.00047195208026096225
step: 190, loss: 0.0001742890162859112
step: 200, loss: 6.573504651896656e-05
step: 210, loss: 5.0410373660270125e-05
step: 220, loss: 0.00046332916826941073
step: 230, loss: 0.0033522718586027622
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```

```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 10: dev_f1=0.9698996655518396, f1=0.966740576496674,
best_f1=0.966740576496674
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.549524838803336e-05
step: 10, loss: 0.00012021389557048678
step: 20, loss: 7.261460996232927e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 30, loss: 8.215041452785954e-05
step: 40, loss: 0.0006890175864100456
step: 50, loss: 0.0002920254191849381
step: 60, loss: 7.556393393315375e-05
step: 70, loss: 0.00023475730267819017
step: 80, loss: 0.00029422654188238084
step: 90, loss: 0.00010252527863485739
step: 100, loss: 6.8520093918778e-05
step: 110, loss: 0.0027095861732959747
step: 120, loss: 0.0007608157466165721
step: 130, loss: 8.46671755425632e-05
step: 140, loss: 4.7914101742208004e-05
step: 150, loss: 4.2376599594717845e-05
step: 160, loss: 0.002854574006050825
step: 170, loss: 0.007338379975408316
step: 180, loss: 0.00010478086915099993
step: 190, loss: 0.0003031285887118429
step: 200, loss: 9.283462713938206e-05
step: 210, loss: 0.00035704439505934715
step: 220, loss: 4.0846174670150504e-05
step: 230, loss: 0.0002961721329484135
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 11: dev_f1=0.9710467706013363, f1=0.9656699889258028,
best_f1=0.9656699889258028
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004417504242155701
step: 10, loss: 0.00011885657295351848

```

```

step: 20, loss: 4.219394031679258e-05
step: 30, loss: 6.33097515674308e-05
step: 40, loss: 4.1601404518587515e-05
step: 50, loss: 9.960975148715079e-05
step: 60, loss: 5.80362684559077e-05
step: 70, loss: 5.5874534155009314e-05
step: 80, loss: 3.169760020682588e-05
step: 90, loss: 7.44468707125634e-05
step: 100, loss: 3.821573045570403e-05
step: 110, loss: 4.450398773769848e-05
step: 120, loss: 4.1278057324234396e-05
step: 130, loss: 8.978947880677879e-05
step: 140, loss: 5.599914220510982e-05
step: 150, loss: 4.59183611383196e-05
step: 160, loss: 5.1356102630961686e-05
step: 170, loss: 8.154183888109401e-05
step: 180, loss: 5.956335735390894e-05
step: 190, loss: 5.90062809351366e-05
step: 200, loss: 0.01293869223445654
step: 210, loss: 0.0008467882871627808
step: 220, loss: 0.04231531545519829
step: 230, loss: 0.0001253935188287869
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9653631284916202, f1=0.9668874172185431,
best_f1=0.9656699889258028
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00014868000289425254
step: 10, loss: 0.00014736127923242748
step: 20, loss: 0.0010022729402408004
step: 30, loss: 0.023998277261853218
step: 40, loss: 0.0006424313760362566
step: 50, loss: 0.0007888735271990299
step: 60, loss: 7.404716598102823e-05
step: 70, loss: 8.41699365992099e-05
step: 80, loss: 4.880533742834814e-05
step: 90, loss: 3.711634053615853e-05
step: 100, loss: 0.034094080328941345
step: 110, loss: 5.5380529374815524e-05
step: 120, loss: 0.00010101980296894908
step: 130, loss: 0.00010599099186947569
step: 140, loss: 0.0006818777765147388
step: 150, loss: 0.044190533459186554
step: 160, loss: 6.993698480073363e-05
step: 170, loss: 0.00018189757247455418

```

```

step: 180, loss: 0.00031439619488082826
step: 190, loss: 4.910663119517267e-05
step: 200, loss: 5.206703644944355e-05
step: 210, loss: 0.0001328154030488804
step: 220, loss: 9.876777039607987e-05
step: 230, loss: 5.353769665816799e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9663677130044843, f1=0.9699666295884317,
best_f1=0.9656699889258028
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.118682747706771e-05
step: 10, loss: 3.220316648366861e-05
step: 20, loss: 0.014906543307006359
step: 30, loss: 0.00010707444744184613
step: 40, loss: 3.869552892865613e-05
step: 50, loss: 0.00011229976371396333
step: 60, loss: 3.119440225418657e-05
step: 70, loss: 5.3166495490586385e-05
step: 80, loss: 0.00012803224672097713
step: 90, loss: 0.0004365125496406108
step: 100, loss: 0.0017766361124813557
step: 110, loss: 0.00011890291352756321
step: 120, loss: 0.0001534083130536601
step: 130, loss: 7.557163189630955e-05
step: 140, loss: 0.0005304593942128122
step: 150, loss: 0.00013990935985930264
step: 160, loss: 3.4494099963922054e-05
step: 170, loss: 4.7426776291104034e-05
step: 180, loss: 5.151754157850519e-05
step: 190, loss: 0.0017850521253421903
step: 200, loss: 3.802619903581217e-05
step: 210, loss: 0.019133325666189194
step: 220, loss: 0.00010688777547329664
step: 230, loss: 3.0076742405071855e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9683972911963882, f1=0.9718785151856018,
best_f1=0.9656699889258028
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.3869400542462245e-05
step: 10, loss: 5.710044570150785e-05

```

```

step: 20, loss: 6.078514707041904e-05
step: 30, loss: 8.820748917059973e-05
step: 40, loss: 8.679658640176058e-05
step: 50, loss: 9.286269778385758e-05
step: 60, loss: 6.359973485814407e-05
step: 70, loss: 4.7793088015168905e-05
step: 80, loss: 0.00045729600242339075
step: 90, loss: 3.2817853934830055e-05
step: 100, loss: 0.00013600377133116126
step: 110, loss: 5.3141251555643976e-05
step: 120, loss: 0.0004286321345716715
step: 130, loss: 7.774640107527375e-05
step: 140, loss: 0.016147945076227188
step: 150, loss: 0.00010647482849890366
step: 160, loss: 4.7514698962913826e-05
step: 170, loss: 0.00013950862921774387
step: 180, loss: 5.265902655082755e-05
step: 190, loss: 0.0001780768361641094
step: 200, loss: 7.480639033019543e-05
step: 210, loss: 0.00934118777513504
step: 220, loss: 0.0027776763308793306
step: 230, loss: 7.312349043786526e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9683257918552037, f1=0.9718785151856018,
best_f1=0.9656699889258028
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic

```

```
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2473it [00:10, 234.24it/s]
load_f1 = 0.9710467706013363
real_f1 = 0.9699666295884317
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:18, 243.61it/s]
```

2.3.2 DBLP-GoogleScholar - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
    --task Dirty/DBLP-GoogleScholar \
    --batch_size 32 \
    --max_len 256 \
    --lr 3e-5 \
    --n_epochs 15 \
    --finetuning \
    --lm distilbert \
    --fp16 \
    --save_model

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-GoogleScholar/.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
    --task Dirty/DBLP-GoogleScholar \
    --input_path input/input_small.jsonl \
    --output_path output/output_small.jsonl \
    --lm distilbert \
    --use_gpu \
    --fp16 \
    --checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:

UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`. In PyTorch 1.1.0 and later, you should call them in the opposite order: `optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

```
"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
```

```
step: 0, loss: 0.7933015823364258
step: 10, loss: 0.4087674617767334
step: 20, loss: 0.5032392740249634
step: 30, loss: 0.42100104689598083
step: 40, loss: 0.3191337585449219
step: 50, loss: 0.18265078961849213
step: 60, loss: 0.2679804563522339
step: 70, loss: 0.15093226730823517
step: 80, loss: 0.2181798219680786
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 90, loss: 0.10528422147035599
step: 100, loss: 0.32940226793289185
```

```
step: 110, loss: 0.028509484604001045
step: 120, loss: 0.03876359015703201
step: 130, loss: 0.07927428930997849
step: 140, loss: 0.14216382801532745
step: 150, loss: 0.03772863373160362
step: 160, loss: 0.18362514674663544
step: 170, loss: 0.16170717775821686
step: 180, loss: 0.15542855858802795
step: 190, loss: 0.055291373282670975
step: 200, loss: 0.08953426778316498
step: 210, loss: 0.1107833981513977
step: 220, loss: 0.06368065625429153
step: 230, loss: 0.12444417178630829
step: 240, loss: 0.09262620657682419
step: 250, loss: 0.061521343886852264
step: 260, loss: 0.06427060812711716
step: 270, loss: 0.013598445802927017
step: 280, loss: 0.12160532921552658
step: 290, loss: 0.08845995366573334
step: 300, loss: 0.08284063637256622
step: 310, loss: 0.06363975256681442
step: 320, loss: 0.021561559289693832
step: 330, loss: 0.09840210527181625
step: 340, loss: 0.13594771921634674
step: 350, loss: 0.15791180729866028
step: 360, loss: 0.07541955262422562
step: 370, loss: 0.13764186203479767
step: 380, loss: 0.263114333152771
step: 390, loss: 0.029353968799114227
step: 400, loss: 0.030267860740423203
step: 410, loss: 0.07302319258451462
step: 420, loss: 0.006627405993640423
step: 430, loss: 0.11857184022665024
step: 440, loss: 0.09618064016103745
step: 450, loss: 0.03555333614349365
step: 460, loss: 0.14655867218971252
step: 470, loss: 0.1556870937347412
step: 480, loss: 0.16645674407482147
step: 490, loss: 0.03679978847503662
step: 500, loss: 0.011384300887584686
step: 510, loss: 0.02726334147155285
step: 520, loss: 0.09528946131467819
step: 530, loss: 0.09710276871919632
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
```



```

loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9314312011044639, f1=0.9307479224376731,
best_f1=0.9307479224376731
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11924247443675995
step: 10, loss: 0.14307881891727448
step: 20, loss: 0.10392417758703232
step: 30, loss: 0.018461808562278748
step: 40, loss: 0.008245809003710747
step: 50, loss: 0.07478699088096619
step: 60, loss: 0.12018640339374542
step: 70, loss: 0.19267906248569489
step: 80, loss: 0.03000130131840706
step: 90, loss: 0.004553909879177809
step: 100, loss: 0.36026573181152344
step: 110, loss: 0.04811437055468559
step: 120, loss: 0.03820467367768288
step: 130, loss: 0.009234598837792873
step: 140, loss: 0.022923197597265244
step: 150, loss: 0.12849903106689453
step: 160, loss: 0.02080201543867588
step: 170, loss: 0.16055196523666382
step: 180, loss: 0.001849092892371118
step: 190, loss: 0.024484053254127502
step: 200, loss: 0.012430093251168728
step: 210, loss: 0.009170813485980034
step: 220, loss: 0.17718946933746338
step: 230, loss: 0.07896364480257034
step: 240, loss: 0.1146693080663681
step: 250, loss: 0.032364822924137115
step: 260, loss: 0.010658566839993
step: 270, loss: 0.11595799028873444
step: 280, loss: 0.12658853828907013
step: 290, loss: 0.11272429674863815
step: 300, loss: 0.04257597774267197
step: 310, loss: 0.03071683645248413
step: 320, loss: 0.11243166029453278
step: 330, loss: 0.03965306282043457
step: 340, loss: 0.0024015074595808983
step: 350, loss: 0.02398582734167576
step: 360, loss: 0.01534238364547491
step: 370, loss: 0.017479581758379936
step: 380, loss: 0.06968733668327332
step: 390, loss: 0.03878556191921234
step: 400, loss: 0.01436595432460308

```

```

step: 410, loss: 0.0004582892288453877
step: 420, loss: 0.07982286810874939
step: 430, loss: 0.020806826651096344
step: 440, loss: 0.02756296843290329
step: 450, loss: 0.01366229634732008
step: 460, loss: 0.20757557451725006
step: 470, loss: 0.055419277399778366
step: 480, loss: 0.1839085966348648
step: 490, loss: 0.037237972021102905
step: 500, loss: 0.0374758243560791
step: 510, loss: 0.05154625326395035
step: 520, loss: 0.07241059094667435
step: 530, loss: 0.10132209956645966
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9381682938168294, f1=0.9312119794103885,
best_f1=0.9312119794103885
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004417069256305695
step: 10, loss: 0.04757055267691612
step: 20, loss: 0.22370116412639618
step: 30, loss: 0.28593340516090393
step: 40, loss: 0.003661587368696928
step: 50, loss: 0.011003787629306316
step: 60, loss: 0.0074509079568088055
step: 70, loss: 0.0074152094312012196
step: 80, loss: 0.0027774351183325052
step: 90, loss: 0.08310116827487946
step: 100, loss: 0.022794397547841072
step: 110, loss: 0.09247341006994247
step: 120, loss: 0.01859704591333866
step: 130, loss: 0.011943896301090717
step: 140, loss: 0.025724561884999275
step: 150, loss: 0.004029769450426102
step: 160, loss: 0.0011806824477389455
step: 170, loss: 0.004250125493854284
step: 180, loss: 0.002899361541494727
step: 190, loss: 0.00787427555769682
step: 200, loss: 0.003262914717197418
step: 210, loss: 0.04348701611161232
step: 220, loss: 0.009307513944804668

```

```

step: 230, loss: 0.015184596180915833
step: 240, loss: 0.017750034108757973
step: 250, loss: 0.005075402092188597
step: 260, loss: 0.0015647687250748277
step: 270, loss: 0.0022111861035227776
step: 280, loss: 0.014474960044026375
step: 290, loss: 0.0495389886200428
step: 300, loss: 0.20133373141288757
step: 310, loss: 0.07501328736543655
step: 320, loss: 0.11038299649953842
step: 330, loss: 0.0016552838496863842
step: 340, loss: 0.004260285757482052
step: 350, loss: 0.07759333401918411
step: 360, loss: 0.006252479273825884
step: 370, loss: 0.001129860756918788
step: 380, loss: 0.0025374717079102993
step: 390, loss: 0.01467120461165905
step: 400, loss: 0.044456981122493744
step: 410, loss: 0.015013977885246277
step: 420, loss: 0.06799064576625824
step: 430, loss: 0.09963031858205795
step: 440, loss: 0.028412340208888054
step: 450, loss: 0.04121146351099014
step: 460, loss: 0.05863259732723236
step: 470, loss: 0.012563350610435009
step: 480, loss: 0.15049518644809723
step: 490, loss: 0.023104721680283546
step: 500, loss: 0.007238004822283983
step: 510, loss: 0.007889891043305397
step: 520, loss: 0.006352601572871208
step: 530, loss: 0.04333186522126198
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.9332711152589829, f1=0.924741298212606,
best_f1=0.9312119794103885
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003120441222563386
step: 10, loss: 0.0033372410107403994
step: 20, loss: 0.015224342234432697
step: 30, loss: 0.0010072411969304085
step: 40, loss: 0.0009446803014725447
step: 50, loss: 0.015642622485756874
step: 60, loss: 0.0003278600051999092
step: 70, loss: 0.000909837952349335
step: 80, loss: 0.010933242738246918

```

```
step: 90, loss: 0.007860051468014717
step: 100, loss: 0.01875692419707775
step: 110, loss: 0.00852280668914318
step: 120, loss: 0.00026179442647844553
step: 130, loss: 0.014561631716787815
step: 140, loss: 0.007153062149882317
step: 150, loss: 0.0019675048533827066
step: 160, loss: 0.003512283321470022
step: 170, loss: 0.0026377830654382706
step: 180, loss: 0.001190914073958993
step: 190, loss: 0.001020199852064252
step: 200, loss: 0.010380418971180916
step: 210, loss: 0.004759638570249081
step: 220, loss: 0.0116723058745265
step: 230, loss: 0.19540685415267944
step: 240, loss: 0.0027940194122493267
step: 250, loss: 0.0008613747777417302
step: 260, loss: 0.07701551914215088
step: 270, loss: 0.010990362614393234
step: 280, loss: 0.0019354004180058837
step: 290, loss: 0.07004108279943466
step: 300, loss: 0.0011828385759145021
step: 310, loss: 0.00613801646977663
step: 320, loss: 0.006289370357990265
step: 330, loss: 0.0035448623821139336
step: 340, loss: 0.004956303630024195
step: 350, loss: 0.001992499688640237
step: 360, loss: 0.007283402141183615
step: 370, loss: 0.01210293173789978
step: 380, loss: 0.008224387653172016
step: 390, loss: 0.03331122174859047
step: 400, loss: 0.006973723415285349
step: 410, loss: 0.0019450021209195256
step: 420, loss: 0.024048512801527977
step: 430, loss: 0.03921472653746605
step: 440, loss: 0.03434067964553833
step: 450, loss: 0.03722207993268967
step: 460, loss: 0.0014714159769937396
step: 470, loss: 0.0030230735428631306
step: 480, loss: 0.0003409363853279501
step: 490, loss: 0.01524405088275671
step: 500, loss: 0.07145389914512634
step: 510, loss: 0.021957451477646828
step: 520, loss: 0.00860767625272274
step: 530, loss: 0.009926891885697842
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.9413970932958275, f1=0.9284020862968232,
best_f1=0.9284020862968232
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0023837543558329344
step: 10, loss: 0.0030657388269901276
step: 20, loss: 0.0006480113952420652
step: 30, loss: 0.0007350666564889252
step: 40, loss: 0.07642275094985962
step: 50, loss: 0.005063147284090519
step: 60, loss: 0.004362727049738169
step: 70, loss: 0.0002257106389151886
step: 80, loss: 0.0004544905386865139
step: 90, loss: 0.0033468434121459723
step: 100, loss: 0.0008044284186325967
step: 110, loss: 0.0018009280320256948
step: 120, loss: 0.0047027613036334515
step: 130, loss: 0.003051079111173749
step: 140, loss: 0.03729458525776863
step: 150, loss: 0.00040126460953615606
step: 160, loss: 0.08567127585411072
step: 170, loss: 0.004239889793097973
step: 180, loss: 0.0012690202565863729
step: 190, loss: 0.003898782655596733
step: 200, loss: 0.0004666569584514946
step: 210, loss: 0.0008627664064988494
step: 220, loss: 0.00016602524556219578
step: 230, loss: 0.0005830675945617259
step: 240, loss: 0.0019179551163688302
step: 250, loss: 0.0004544538969639689
step: 260, loss: 0.02123943157494068
step: 270, loss: 0.008595052175223827
step: 280, loss: 0.05899541452527046
step: 290, loss: 0.00019726586469914764
step: 300, loss: 0.019443903118371964
step: 310, loss: 0.0011835901532322168
step: 320, loss: 0.006170580163598061
step: 330, loss: 0.0014929900644347072
step: 340, loss: 0.00324849016033113
step: 350, loss: 0.02961411327123642
step: 360, loss: 0.005097472108900547
step: 370, loss: 0.011269529350101948
step: 380, loss: 0.027786800637841225

```

```

step: 390, loss: 0.002132045105099678
step: 400, loss: 0.00033589918166399
step: 410, loss: 0.0016799408476799726
step: 420, loss: 0.002260254230350256
step: 430, loss: 0.013307753019034863
step: 440, loss: 0.001294823014177382
step: 450, loss: 0.0003885991172865033
step: 460, loss: 0.0002918084501288831
step: 470, loss: 0.0020014611072838306
step: 480, loss: 0.010193810798227787
step: 490, loss: 0.0014947172021493316
step: 500, loss: 0.009786025620996952
step: 510, loss: 0.0013085651444271207
step: 520, loss: 0.000860001367982477
step: 530, loss: 0.007177822757512331
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.9401789919924635, f1=0.927315914489311,
best_f1=0.9284020862968232
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006096964702010155
step: 10, loss: 0.0005121629219502211
step: 20, loss: 0.0003263523685745895
step: 30, loss: 0.01151775848120451
step: 40, loss: 0.001111573539674282
step: 50, loss: 0.11875496059656143
step: 60, loss: 0.002595315920189023
step: 70, loss: 0.000715186120942235
step: 80, loss: 0.00988711230456829
step: 90, loss: 0.0003792629577219486
step: 100, loss: 0.047690510749816895
step: 110, loss: 0.004899085033684969
step: 120, loss: 0.0005220607272349298
step: 130, loss: 0.00022672036720905453
step: 140, loss: 0.0053041474893689156
step: 150, loss: 0.00018314139742869884
step: 160, loss: 0.00013237916573416442
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 170, loss: 0.00013039047189522535
step: 180, loss: 0.011826563626527786
step: 190, loss: 0.00040881408494897187
step: 200, loss: 0.0002747262769844383
step: 210, loss: 0.0010981714585795999
step: 220, loss: 0.008208858780562878
step: 230, loss: 0.0001399400207446888

```

```

step: 240, loss: 0.001479326980188489
step: 250, loss: 0.0006256335764192045
step: 260, loss: 0.000849218457005918
step: 270, loss: 0.004106537904590368
step: 280, loss: 0.0013955957256257534
step: 290, loss: 0.0008779017371125519
step: 300, loss: 0.0013568782014772296
step: 310, loss: 0.001105405855923891
step: 320, loss: 0.16355057060718536
step: 330, loss: 0.018184026703238487
step: 340, loss: 0.003082670271396637
step: 350, loss: 0.012477093376219273
step: 360, loss: 0.0005069300532341003
step: 370, loss: 0.007306419778615236
step: 380, loss: 0.0011212073732167482
step: 390, loss: 0.0014944347785785794
step: 400, loss: 0.0013265921734273434
step: 410, loss: 0.00022822168830316514
step: 420, loss: 0.009771827608346939
step: 430, loss: 0.0005300156190060079
step: 440, loss: 0.0010888951364904642
step: 450, loss: 0.0009089953964576125
step: 460, loss: 0.0005882284604012966
step: 470, loss: 0.011167098768055439
step: 480, loss: 0.008058052510023117
step: 490, loss: 0.00029801978962495923
step: 500, loss: 0.006132963113486767
step: 510, loss: 0.00021496054250746965
step: 520, loss: 0.00896795466542244
step: 530, loss: 0.001450766809284687
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 6: dev_f1=0.9433085501858736, f1=0.929840972871843,
best_f1=0.929840972871843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001197130186483264
step: 10, loss: 0.00710391066968441
step: 20, loss: 0.0006889637443237007
step: 30, loss: 0.020417222753167152
step: 40, loss: 0.00015418381372001022
step: 50, loss: 0.0006350017501972616

```

step: 60, loss: 0.0002784082025755197
step: 70, loss: 0.0020091517362743616
step: 80, loss: 0.06663238257169724
step: 90, loss: 0.00013876633602194488
step: 100, loss: 0.0005553559167310596
step: 110, loss: 0.00025514408480376005
step: 120, loss: 0.0009661635267548263
step: 130, loss: 6.256917549762875e-05
step: 140, loss: 0.005107786972075701
step: 150, loss: 8.153216913342476e-05
step: 160, loss: 0.0007478270563296974
step: 170, loss: 0.0004139191296417266
step: 180, loss: 0.00044994152267463505
step: 190, loss: 9.057485294761136e-05
step: 200, loss: 0.00013819537707604468
step: 210, loss: 8.059607353061438e-05
step: 220, loss: 0.000561786990147084
step: 230, loss: 0.00031796962139196694
step: 240, loss: 0.001535782590508461
step: 250, loss: 8.857215289026499e-05
step: 260, loss: 0.000249047385295853
step: 270, loss: 0.00017143102013505995
step: 280, loss: 0.04204052686691284
step: 290, loss: 0.06349413096904755
step: 300, loss: 0.0028478512540459633
step: 310, loss: 0.0001568353909533471
step: 320, loss: 0.026501083746552467
step: 330, loss: 0.000804430921562016
step: 340, loss: 0.0031646371353417635
step: 350, loss: 0.008165544830262661
step: 360, loss: 0.003619384253397584
step: 370, loss: 0.007518216967582703
step: 380, loss: 0.0020107056479901075
step: 390, loss: 0.00011313988943584263
step: 400, loss: 0.00028323029982857406
step: 410, loss: 0.0007883953512646258
step: 420, loss: 0.000275561265880242
step: 430, loss: 5.426934512797743e-05
step: 440, loss: 0.00017032363393809646
step: 450, loss: 0.0037145758979022503
step: 460, loss: 0.011286898516118526
step: 470, loss: 0.0075388322584331036
step: 480, loss: 0.0015780874527990818
step: 490, loss: 0.002715132897719741
step: 500, loss: 0.0010544423712417483
step: 510, loss: 0.0010268895421177149
step: 520, loss: 0.0008326285169459879
step: 530, loss: 8.661359606776386e-05


```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9353468617272298, f1=0.925627664613927,
best_f1=0.929840972871843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0032499961089342833
step: 10, loss: 0.014955266378819942
step: 20, loss: 0.012106089852750301
step: 30, loss: 0.0017591570504009724
step: 40, loss: 7.954531611176208e-05
step: 50, loss: 0.0002351626317249611
step: 60, loss: 0.0007102183299139142
step: 70, loss: 0.00011112799256807193
step: 80, loss: 0.00020764186047017574
step: 90, loss: 0.001926324563100934
step: 100, loss: 0.0008894939674064517
step: 110, loss: 0.0007309525390155613
step: 120, loss: 0.00038910386501811445
step: 130, loss: 0.00025995654868893325
step: 140, loss: 0.00015631102724000812
step: 150, loss: 0.0020917546935379505
step: 160, loss: 0.00030968867940828204
step: 170, loss: 0.007203764747828245
step: 180, loss: 0.00048271004925481975
step: 190, loss: 0.004521416500210762
step: 200, loss: 0.0004868621181230992
step: 210, loss: 0.0007361036259680986
step: 220, loss: 0.0006368323811329901
step: 230, loss: 0.0002842347021214664
step: 240, loss: 7.940611249068752e-05
step: 250, loss: 0.0015443511074408889
step: 260, loss: 0.00581471947953105
step: 270, loss: 4.838130917050876e-05
step: 280, loss: 0.0011563501320779324
step: 290, loss: 0.00010039285552920774
step: 300, loss: 6.801920244470239e-05
step: 310, loss: 0.014656202867627144
step: 320, loss: 0.022104430943727493
step: 330, loss: 0.0037785894237458706
step: 340, loss: 8.798404451226816e-05
step: 350, loss: 0.08359748125076294
step: 360, loss: 0.0010141204111278057
step: 370, loss: 0.001008707331493497
step: 380, loss: 0.013611311092972755
step: 390, loss: 0.00014837108028586954
```

```

step: 400, loss: 0.08976172655820847
step: 410, loss: 0.0001452486903872341
step: 420, loss: 6.061475505703129e-05
step: 430, loss: 0.0022888092789798975
step: 440, loss: 0.0008626498165540397
step: 450, loss: 0.003950740210711956
step: 460, loss: 0.023011978715658188
step: 470, loss: 0.0001250366767635569
step: 480, loss: 0.0015484426403418183
step: 490, loss: 0.0005124175804667175
step: 500, loss: 0.0006908017676323652
step: 510, loss: 0.00011580433783819899
step: 520, loss: 0.00011553162039490417
step: 530, loss: 0.0005738871986977756
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9394495412844037, f1=0.9287037037037037,
best_f1=0.929840972871843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001846809609560296
step: 10, loss: 0.0003023750032298267
step: 20, loss: 0.00026628479827195406
step: 30, loss: 0.00019971156143583357
step: 40, loss: 0.0008990904898382723
step: 50, loss: 0.00031951910932548344
step: 60, loss: 7.795167766744271e-05
step: 70, loss: 0.1274998039007187
step: 80, loss: 0.0005553033552132547
step: 90, loss: 0.00026570583577267826
step: 100, loss: 0.0004718565905932337
step: 110, loss: 0.0004414712602738291
step: 120, loss: 6.0248632507864386e-05
step: 130, loss: 0.00022569346765521914
step: 140, loss: 0.002290197415277362
step: 150, loss: 0.0003548417880665511
step: 160, loss: 6.727071013301611e-05
step: 170, loss: 0.0003745517460629344
step: 180, loss: 0.0008339565829373896
step: 190, loss: 0.00044989914749749005
step: 200, loss: 0.00037193583557382226
step: 210, loss: 4.1028881241800264e-05
step: 220, loss: 0.0012453830568119884
step: 230, loss: 6.309105810942128e-05
step: 240, loss: 0.000376547162886709
step: 250, loss: 0.0006401588907465339

```

```

step: 260, loss: 0.0010662027634680271
step: 270, loss: 0.00020543414575513452
step: 280, loss: 5.062328273197636e-05
step: 290, loss: 5.7454759371466935e-05
step: 300, loss: 0.00018056279805023223
step: 310, loss: 8.651458483655006e-05
step: 320, loss: 0.016136417165398598
step: 330, loss: 0.03720290958881378
step: 340, loss: 0.00047819671453908086
step: 350, loss: 6.844590825494379e-05
step: 360, loss: 0.01787494122982025
step: 370, loss: 0.0007296483963727951
step: 380, loss: 4.364687993074767e-05
step: 390, loss: 8.251193503383547e-05
step: 400, loss: 0.009977156296372414
step: 410, loss: 8.021032408578321e-05
step: 420, loss: 0.00018390057084616274
step: 430, loss: 6.23452287982218e-05
step: 440, loss: 6.376896635629237e-05
step: 450, loss: 9.063630568562075e-05
step: 460, loss: 0.0002649183734320104
step: 470, loss: 0.011629697866737843
step: 480, loss: 0.00033316542976535857
step: 490, loss: 7.872482092352584e-05
step: 500, loss: 0.0009758528904058039
step: 510, loss: 0.0007846967200748622
step: 520, loss: 0.00011824337707366794
step: 530, loss: 8.910788892535493e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9391955098222639, f1=0.9284712482468443,
best_f1=0.929840972871843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05528199300169945
step: 10, loss: 2.6963005439029075e-05
step: 20, loss: 0.032156314700841904
step: 30, loss: 3.875933543895371e-05
step: 40, loss: 0.003089467529207468
step: 50, loss: 0.0006065224879421294
step: 60, loss: 5.563552986131981e-05
step: 70, loss: 0.0006205656100064516
step: 80, loss: 0.0006304422277025878
step: 90, loss: 4.6757606469327584e-05
step: 100, loss: 0.0002306034293724224
step: 110, loss: 0.00043345033191144466

```

```

step: 120, loss: 0.00013458759349305183
step: 130, loss: 3.2386778912041336e-05
step: 140, loss: 0.0005103231524117291
step: 150, loss: 3.146269591525197e-05
step: 160, loss: 3.8234829844441265e-05
step: 170, loss: 0.0008081815321929753
step: 180, loss: 0.00032505678245797753
step: 190, loss: 0.000804462528321892
step: 200, loss: 0.0006641779327765107
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 210, loss: 9.354404755868018e-05
step: 220, loss: 0.0001747888745740056
step: 230, loss: 4.839439861825667e-05
step: 240, loss: 0.00019431623513810337
step: 250, loss: 3.069901868002489e-05
step: 260, loss: 0.003071446903049946
step: 270, loss: 3.0118124414002523e-05
step: 280, loss: 0.0005269173416309059
step: 290, loss: 5.638736911350861e-05
step: 300, loss: 8.042575791478157e-05
step: 310, loss: 0.0012358974199742079
step: 320, loss: 9.12710020202212e-05
step: 330, loss: 3.502416439005174e-05
step: 340, loss: 3.674844265333377e-05
step: 350, loss: 3.33329699060414e-05
step: 360, loss: 0.00020129459153395146
step: 370, loss: 0.0003241637896280736
step: 380, loss: 9.058167779585347e-05
step: 390, loss: 3.484144326648675e-05
step: 400, loss: 7.548182475147769e-05
step: 410, loss: 0.00020944321295246482
step: 420, loss: 5.741826316807419e-05
step: 430, loss: 6.013684833305888e-05
step: 440, loss: 3.343314165249467e-05
step: 450, loss: 0.00021347569418139756
step: 460, loss: 0.0003339366230648011
step: 470, loss: 3.381706846994348e-05
step: 480, loss: 0.0002935734810307622
step: 490, loss: 0.035114966332912445
step: 500, loss: 0.00557286711409688
step: 510, loss: 0.0026323620695620775
step: 520, loss: 0.00017876445781439543
step: 530, loss: 3.779678445425816e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9314045730284647, f1=0.9231494578029231,
best_f1=0.929840972871843

```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:  
An input tensor was not cuda.
```

```
    warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 5.8057605201611295e-05  
step: 10, loss: 0.00846659205853939  
step: 20, loss: 3.868983549182303e-05  
step: 30, loss: 9.104268247028813e-05  
step: 40, loss: 7.845111394999549e-05  
step: 50, loss: 0.0002614050463307649  
step: 60, loss: 3.075103450100869e-05  
step: 70, loss: 3.435686812736094e-05  
step: 80, loss: 0.0014846554258838296  
step: 90, loss: 0.00012980405881535262  
step: 100, loss: 5.802645318908617e-05  
step: 110, loss: 7.121271482901648e-05  
step: 120, loss: 6.611777644138783e-05  
step: 130, loss: 0.0005651141400448978  
step: 140, loss: 5.6075165048241615e-05  
step: 150, loss: 3.623395969043486e-05  
step: 160, loss: 0.0006013347883708775  
step: 170, loss: 8.865413838066161e-05  
step: 180, loss: 3.0360319215105847e-05  
step: 190, loss: 0.0003197408514097333  
step: 200, loss: 0.0005523486761376262  
step: 210, loss: 4.3456991988932714e-05  
step: 220, loss: 3.982185808126815e-05  
step: 230, loss: 5.960316048003733e-05  
step: 240, loss: 3.6565314076142386e-05  
step: 250, loss: 3.122726047877222e-05  
step: 260, loss: 0.017561161890625954  
step: 270, loss: 0.0006284157279878855  
step: 280, loss: 5.0518487114459276e-05  
step: 290, loss: 0.0031706183217465878  
step: 300, loss: 0.00026288742083124816  
step: 310, loss: 5.7805886171991006e-05  
step: 320, loss: 0.006526874378323555  
step: 330, loss: 0.0003143135691061616  
step: 340, loss: 0.00018300663214176893  
step: 350, loss: 0.00028285401640459895  
step: 360, loss: 6.043647226761095e-05  
step: 370, loss: 2.2783315216656774e-05  
step: 380, loss: 6.8301145802252e-05  
step: 390, loss: 0.003184621687978506  
step: 400, loss: 3.9763541280990466e-05  
step: 410, loss: 0.00011478487431304529  
step: 420, loss: 0.0016690552001819015  
step: 430, loss: 0.0011099587427452207  
step: 440, loss: 8.473184425383806e-05
```

```

step: 450, loss: 3.2337757147615775e-05
step: 460, loss: 0.0027371207252144814
step: 470, loss: 0.00019479036564007401
step: 480, loss: 8.361263462575153e-05
step: 490, loss: 0.00035622797440737486
step: 500, loss: 3.615561217884533e-05
step: 510, loss: 3.0374809284694493e-05
step: 520, loss: 2.524551200622227e-05
step: 530, loss: 3.0049866836634465e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9243295019157088, f1=0.916866890917828,
best_f1=0.929840972871843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0042340620420873165
step: 10, loss: 0.01470867171883583
step: 20, loss: 2.948827932414133e-05
step: 30, loss: 6.219376518856734e-05
step: 40, loss: 4.171156251686625e-05
step: 50, loss: 7.728579657850787e-05
step: 60, loss: 3.728805677383207e-05
step: 70, loss: 5.7964301959145814e-05
step: 80, loss: 0.00010226402810076252
step: 90, loss: 2.711200068006292e-05
step: 100, loss: 2.3725842765998095e-05
step: 110, loss: 4.67285790364258e-05
step: 120, loss: 6.884751928737387e-05
step: 130, loss: 3.720237145898864e-05
step: 140, loss: 6.500452582258731e-05
step: 150, loss: 5.3055940952617675e-05
step: 160, loss: 0.0001587603474035859
step: 170, loss: 2.2805805201642215e-05
step: 180, loss: 0.0003897623100783676
step: 190, loss: 0.0001066140248440206
step: 200, loss: 3.921585448551923e-05
step: 210, loss: 0.00016464079089928418
step: 220, loss: 2.632961331983097e-05
step: 230, loss: 0.0055322046391665936
step: 240, loss: 0.00013863903586752713
step: 250, loss: 2.7238505936111324e-05
step: 260, loss: 0.0005815651384182274
step: 270, loss: 0.0005437061190605164
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 280, loss: 2.685860454221256e-05
step: 290, loss: 0.004765330348163843

```

```

step: 300, loss: 4.652963616535999e-05
step: 310, loss: 4.9596186727285385e-05
step: 320, loss: 3.072905383305624e-05
step: 330, loss: 2.8668819140875712e-05
step: 340, loss: 0.0006305602728389204
step: 350, loss: 0.0022088661789894104
step: 360, loss: 7.35919747967273e-05
step: 370, loss: 2.3539365429314785e-05
step: 380, loss: 4.505470860749483e-05
step: 390, loss: 5.0612543418537825e-05
step: 400, loss: 4.634910146705806e-05
step: 410, loss: 0.012427093461155891
step: 420, loss: 0.0001857215684140101
step: 430, loss: 7.705703319516033e-05
step: 440, loss: 0.0002541947760619223
step: 450, loss: 0.00011752339196391404
step: 460, loss: 0.00026249009533785284
step: 470, loss: 4.451662607607432e-05
step: 480, loss: 4.8894118663156405e-05
step: 490, loss: 0.00031173217575997114
step: 500, loss: 0.00010025405936175957
step: 510, loss: 5.0786456995410845e-05
step: 520, loss: 0.0074888113886117935
step: 530, loss: 4.783963231602684e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9381918819188192, f1=0.9326568265682658,
best_f1=0.929840972871843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.0401766455033794e-05
step: 10, loss: 0.002005192684009671
step: 20, loss: 0.0005202192696742713
step: 30, loss: 0.0006939556915313005
step: 40, loss: 3.17496269417461e-05
step: 50, loss: 0.00012222684745211154
step: 60, loss: 5.687263183062896e-05
step: 70, loss: 8.020373206818476e-05
step: 80, loss: 3.124235081486404e-05
step: 90, loss: 0.012383061461150646
step: 100, loss: 4.147800791542977e-05
step: 110, loss: 7.07266153767705e-05
step: 120, loss: 0.0001128416188294068
step: 130, loss: 0.000621287093963474
step: 140, loss: 0.0006785524310544133
step: 150, loss: 5.009284723200835e-05

```

```

step: 160, loss: 2.3718424927210435e-05
step: 170, loss: 7.267644832609221e-05
step: 180, loss: 3.424415990593843e-05
step: 190, loss: 4.2763014789670706e-05
step: 200, loss: 0.006266014184802771
step: 210, loss: 0.0017707212828099728
step: 220, loss: 2.941255115729291e-05
step: 230, loss: 0.0007292255177162588
step: 240, loss: 8.15981620689854e-05
step: 250, loss: 0.051866184920072556
step: 260, loss: 0.028309397399425507
step: 270, loss: 5.131507714395411e-05
step: 280, loss: 0.0009514117846265435
step: 290, loss: 5.643198164761998e-05
step: 300, loss: 0.0001402936177328229
step: 310, loss: 0.0014798152260482311
step: 320, loss: 0.00030605742358602583
step: 330, loss: 8.293266728287563e-05
step: 340, loss: 0.00016195696662180126
step: 350, loss: 0.003489595605060458
step: 360, loss: 9.934788249665871e-05
step: 370, loss: 7.267182809300721e-05
step: 380, loss: 0.00015151232946664095
step: 390, loss: 3.270279194111936e-05
step: 400, loss: 2.514088191674091e-05
step: 410, loss: 3.920188100892119e-05
step: 420, loss: 9.123337076744065e-05
step: 430, loss: 0.01629529520869255
step: 440, loss: 0.0004853567515965551
step: 450, loss: 0.00981017667800188
step: 460, loss: 2.4068487618933432e-05
step: 470, loss: 0.0006595239392481744
step: 480, loss: 0.0006632988806813955
step: 490, loss: 0.0020899970550090075
step: 500, loss: 0.0003348042373545468
step: 510, loss: 2.5472854758845642e-05
step: 520, loss: 3.0769304430577904e-05
step: 530, loss: 0.00010651900811353698
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9379115710253997, f1=0.9342723004694835,
best_f1=0.929840972871843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001669949560891837
step: 10, loss: 7.251611532410607e-05

```


step: 20, loss: 2.0533385395538062e-05
step: 30, loss: 3.274804475950077e-05
step: 40, loss: 5.907305967411958e-05
step: 50, loss: 4.918637569062412e-05
step: 60, loss: 4.0100196201819927e-05
step: 70, loss: 2.6880637960857712e-05
step: 80, loss: 0.00018816177907865494
step: 90, loss: 2.417646646790672e-05
step: 100, loss: 0.00021922760060988367
step: 110, loss: 3.356174420332536e-05
step: 120, loss: 2.494261207175441e-05
step: 130, loss: 2.5960613129427657e-05
step: 140, loss: 5.128423072164878e-05
step: 150, loss: 4.4454020098783076e-05
step: 160, loss: 0.004516131244599819
step: 170, loss: 0.00026650994550436735
step: 180, loss: 0.00015518185682594776
step: 190, loss: 0.0018886057659983635
step: 200, loss: 3.869281499646604e-05
step: 210, loss: 0.00031628692522644997
step: 220, loss: 9.43562772590667e-05
step: 230, loss: 0.0006129933171905577
step: 240, loss: 7.112661842256784e-05
step: 250, loss: 5.780518404208124e-05
step: 260, loss: 2.0198043785057962e-05
step: 270, loss: 0.0022985204122960567
step: 280, loss: 2.532366124796681e-05
step: 290, loss: 6.126469816081226e-05
step: 300, loss: 0.0023832328151911497
step: 310, loss: 2.2905998775968328e-05
step: 320, loss: 7.150052988436073e-05
step: 330, loss: 0.0034460455644875765
step: 340, loss: 4.3295149225741625e-05
step: 350, loss: 0.0001771734532667324
step: 360, loss: 0.0028545644599944353
step: 370, loss: 2.7833930289489217e-05
step: 380, loss: 2.0518442397587933e-05
step: 390, loss: 6.331660551950336e-05
step: 400, loss: 2.9380558771663345e-05
step: 410, loss: 2.628064430609811e-05
step: 420, loss: 6.315071368589997e-05
step: 430, loss: 3.94786948163528e-05
step: 440, loss: 1.5634792362106964e-05
step: 450, loss: 2.3595275706611574e-05
step: 460, loss: 0.0001059033747878857
step: 470, loss: 2.324128035979811e-05
step: 480, loss: 2.0499781385296956e-05
step: 490, loss: 3.6106444895267487e-05

```

step: 500, loss: 0.00016004944336600602
step: 510, loss: 3.0437899113167077e-05
step: 520, loss: 6.332479824777693e-05
step: 530, loss: 0.0001890919666038826
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9371158392434988, f1=0.9291115311909263,
best_f1=0.929840972871843
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.088546095066704e-05
step: 10, loss: 8.012135367607698e-05
step: 20, loss: 2.3878474166849628e-05
step: 30, loss: 5.101375063532032e-05
step: 40, loss: 2.6588320906739682e-05
step: 50, loss: 0.0004340402374509722
step: 60, loss: 2.45520150201628e-05
step: 70, loss: 0.00013035372830927372
step: 80, loss: 7.356407877523452e-05
step: 90, loss: 2.5677385565359145e-05
step: 100, loss: 2.472746928106062e-05
step: 110, loss: 0.0003774996439460665
step: 120, loss: 0.0017140241106972098
step: 130, loss: 1.7240356100955978e-05
step: 140, loss: 1.7299938917858526e-05
step: 150, loss: 7.807368820067495e-05
step: 160, loss: 0.00015708347200416028
step: 170, loss: 4.966762207914144e-05
step: 180, loss: 2.107707769027911e-05
step: 190, loss: 0.0009657446062192321
step: 200, loss: 3.269487206125632e-05
step: 210, loss: 3.2240088330581784e-05
step: 220, loss: 1.3906354979553726e-05
step: 230, loss: 7.710024510743096e-05
step: 240, loss: 5.955469168839045e-05
step: 250, loss: 0.0005332757136784494
step: 260, loss: 0.027609696611762047
step: 270, loss: 0.00020348177349660546
step: 280, loss: 0.0004110177978873253
step: 290, loss: 0.00017652720271144062
step: 300, loss: 5.356776819098741e-05
step: 310, loss: 2.707439307414461e-05
step: 320, loss: 9.032770321937278e-05
step: 330, loss: 4.638915561372414e-05
step: 340, loss: 6.053023025742732e-05
step: 350, loss: 3.076935900026001e-05

```

```

step: 360, loss: 0.0001583173288963735
step: 370, loss: 2.1866679162485525e-05
step: 380, loss: 2.401252459094394e-05
step: 390, loss: 2.839180888258852e-05
step: 400, loss: 4.3245650886092335e-05
step: 410, loss: 5.596744813374244e-05
step: 420, loss: 3.157401806674898e-05
step: 430, loss: 2.0328499886090867e-05
step: 440, loss: 3.732708137249574e-05
step: 450, loss: 0.0003644298994913697
step: 460, loss: 1.701681685517542e-05
step: 470, loss: 8.29357813927345e-05
step: 480, loss: 2.3680919184698723e-05
step: 490, loss: 3.108631062787026e-05
step: 500, loss: 7.750759687041864e-05
step: 510, loss: 7.795581041136757e-05
step: 520, loss: 0.0002158414717996493
step: 530, loss: 6.201241194503382e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9369202226345084, f1=0.932901434521055,
best_f1=0.929840972871843
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
5742it [00:23, 245.74it/s]
load_f1 = 0.9434137291280148
real_f1 = 0.9403054141601109
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:17, 246.65it/s]

```

3 DK

```

[ ]: # Para ejecutar dk necesita
      #!pip install -r requirements.txt
      !python -m spacy download en_core_web_lg

```

```

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Collecting en_core_web_lg==2.2.5
  Downloading https://github.com/explosion/spacy-
models/releases/download/en_core_web_lg-2.2.5/en_core_web_lg-2.2.5.tar.gz (827.9
MB)
    |                               | 827.9 MB 1.2 MB/s
Requirement already satisfied: spacy>=2.2.2 in
/usr/local/lib/python3.7/dist-packages (from en_core_web_lg==2.2.5) (2.2.3)
Requirement already satisfied: srsly<1.1.0,>=0.1.0 in
/usr/local/lib/python3.7/dist-packages (from
spacy>=2.2.2->en_core_web_lg==2.2.5) (1.0.5)
Requirement already satisfied: plac<1.2.0,>=0.9.6 in
/usr/local/lib/python3.7/dist-packages (from
spacy>=2.2.2->en_core_web_lg==2.2.5) (1.1.3)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in
/usr/local/lib/python3.7/dist-packages (from
spacy>=2.2.2->en_core_web_lg==2.2.5) (2.0.6)
Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in
/usr/local/lib/python3.7/dist-packages (from
spacy>=2.2.2->en_core_web_lg==2.2.5) (1.0.0)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in
/usr/local/lib/python3.7/dist-packages (from
spacy>=2.2.2->en_core_web_lg==2.2.5) (2.23.0)
Requirement already satisfied: blis<0.5.0,>=0.4.0 in
/usr/local/lib/python3.7/dist-packages (from
spacy>=2.2.2->en_core_web_lg==2.2.5) (0.4.1)

```

Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from spacy>=2.2.2->en_core_web_lg==2.2.5) (3.0.6)

Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in /usr/local/lib/python3.7/dist-packages (from spacy>=2.2.2->en_core_web_lg==2.2.5) (0.10.1)

Requirement already satisfied: thinc<7.4.0,>=7.3.0 in /usr/local/lib/python3.7/dist-packages (from spacy>=2.2.2->en_core_web_lg==2.2.5) (7.3.1)

Requirement already satisfied: numpy>=1.15.0 in /usr/local/lib/python3.7/dist-packages (from spacy>=2.2.2->en_core_web_lg==2.2.5) (1.19.2)

Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.7/dist-packages (from spacy>=2.2.2->en_core_web_lg==2.2.5) (1.0.7)

Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages (from spacy>=2.2.2->en_core_web_lg==2.2.5) (57.4.0)

Requirement already satisfied: importlib-metadata>=0.20 in /usr/local/lib/python3.7/dist-packages (from catalogue<1.1.0,>=0.0.7->spacy>=2.2.2->en_core_web_lg==2.2.5) (4.12.0)

Requirement already satisfied: typing-extensions>=3.6.4 in /usr/local/lib/python3.7/dist-packages (from importlib-metadata>=0.20->catalogue<1.1.0,>=0.0.7->spacy>=2.2.2->en_core_web_lg==2.2.5) (4.1.1)

Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (from importlib-metadata>=0.20->catalogue<1.1.0,>=0.0.7->spacy>=2.2.2->en_core_web_lg==2.2.5) (3.8.1)

Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests<3.0.0,>=2.13.0->spacy>=2.2.2->en_core_web_lg==2.2.5) (2.10)

Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages (from requests<3.0.0,>=2.13.0->spacy>=2.2.2->en_core_web_lg==2.2.5) (1.25.4)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from requests<3.0.0,>=2.13.0->spacy>=2.2.2->en_core_web_lg==2.2.5) (2022.6.15)

Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from requests<3.0.0,>=2.13.0->spacy>=2.2.2->en_core_web_lg==2.2.5) (3.0.4)

Requirement already satisfied: tqdm<5.0.0,>=4.10.0 in /usr/local/lib/python3.7/dist-packages (from thinc<7.4.0,>=7.3.0->spacy>=2.2.2->en_core_web_lg==2.2.5) (4.41.0)

Building wheels for collected packages: en-core-web-lg

Building wheel for en-core-web-lg (setup.py) ... done

Created wheel for en-core-web-lg: filename=en_core_web_lg-2.2.5-py3-none-any.whl size=829180942 sha256=39db757681decfafa0c7e5e38e43ddedd1452b123e55f887e0a7836830d08a69

Stored in directory: /tmp/pip-ephem-wheel-cache-

```
drnky8bl/wheels/11/95/ba/2c36cc368c0bd339b44a791c2c1881a1fb714b78c29a4cb8f5
Successfully built en-core-web-lg
Installing collected packages: en-core-web-lg
Successfully installed en-core-web-lg-2.2.5
Download and installation successful
You can now load the model via spacy.load('en_core_web_lg')
```

3.1 DK STRUCTURED

3.1.1 Beer - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Structured/Beer \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model \
  --dk general \
  --dk product \
  --summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Beer.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/Beer \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8494251370429993
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.2666666666666667, f1=0.2666666666666667,
best_f1=0.2666666666666667
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.37428444623947144
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.3636363636363636, f1=0.3466666666666666,
best_f1=0.3466666666666666
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.32705989480018616
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.41791044776119407, f1=0.3492063492063492,
best_f1=0.3492063492063492
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3454452157020569
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.7692307692307692, f1=0.6250000000000001,
best_f1=0.6250000000000001
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.37552326917648315
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.7741935483870968, f1=0.5789473684210527,
best_f1=0.5789473684210527
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```



```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.23280496895313263
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.6666666666666666, f1=0.5555555555555556,
best_f1=0.5789473684210527
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1960223913192749
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.8148148148148148, f1=0.7741935483870968,
best_f1=0.7741935483870968
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.19052158296108246
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.717948717948718, f1=0.6341463414634146,
best_f1=0.7741935483870968
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.18901361525058746
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 9: dev_f1=0.9333333333333333, f1=0.8387096774193549,
best_f1=0.8387096774193549
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04317783564329147
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9032258064516129, f1=0.8484848484848484,
best_f1=0.8387096774193549
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.040039755403995514
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9032258064516129, f1=0.8484848484848484,
best_f1=0.8387096774193549
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04311239719390869
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9032258064516129, f1=0.8484848484848484,
best_f1=0.8387096774193549
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.034040667116642
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9032258064516129, f1=0.8484848484848484,
best_f1=0.8387096774193549
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.021898120641708374
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9032258064516129, f1=0.8484848484848484,
best_f1=0.8387096774193549
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.022098219022154808
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9032258064516129, f1=0.8484848484848484,
best_f1=0.8387096774193549

```

```
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2: FP16 training with FP32 batchnorm and FP32
master weights.
```

Defaults for this optimization level are:

```
enabled          : True
opt_level        : O2
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True
opt_level        : O2
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

```
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
warnings.warn("An input tensor was not cuda.")
```

```
91it [00:00, 117948.60it/s]
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
warnings.warn("An input tensor was not cuda.")
```

```
load_f1 = 0.8484848484848484
```

```
real_f1 = 0.8
```

```
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
```

```
warnings.warn("An input tensor was not cuda.")
```

```
4398it [00:10, 422.35it/s]
```

3.1.2 DBLP-ACM - Running the matcher

```
[ ]: # Step 1:
      # train the wdc model with the save_model flag on
      # it will produce two files *_dev.pt and *_test.pt which are the
      # best checkpoints on the dev set and the test set.
      !CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
        --task Structured/DBLP-ACM \
```

```

--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \
--dk product \
--summarize

```

Step 2: pick a checkpoint, rename it to wdc_all_small.pt

```
!mv *_dev.pt checkpoints/DBLP-ACM.pt
```

Step 3: run the matcher

```

!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/DBLP-ACM \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly

```

because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.7738555669784546
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.44801804423332214
step: 20, loss: 0.2994305193424225
step: 30, loss: 0.219679594039917
step: 40, loss: 0.12398871034383774
step: 50, loss: 0.08667023479938507
step: 60, loss: 0.13214559853076935
step: 70, loss: 0.009543520398437977
step: 80, loss: 0.06112760677933693
step: 90, loss: 0.05755511671304703
step: 100, loss: 0.014092080295085907
step: 110, loss: 0.00793736893683672
step: 120, loss: 0.010213745757937431
step: 130, loss: 0.002367389155551791
step: 140, loss: 0.024454481899738312
step: 150, loss: 0.07251708954572678
step: 160, loss: 0.03579094633460045
step: 170, loss: 0.003241908270865679
step: 180, loss: 0.012521850876510143
step: 190, loss: 0.01845083199441433
step: 200, loss: 0.0018245697719976306
step: 210, loss: 0.019467217847704887
step: 220, loss: 0.03838495537638664
step: 230, loss: 0.03540750592947006
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9853768278965129, f1=0.9785794813979707,

```

```

best_f1=0.9785794813979707
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005317847244441509
step: 10, loss: 0.0029259533621370792
step: 20, loss: 0.11176678538322449
step: 30, loss: 0.0023220013827085495
step: 40, loss: 0.0020389449782669544
step: 50, loss: 0.029445424675941467
step: 60, loss: 0.11930706351995468
step: 70, loss: 0.01150591392070055
step: 80, loss: 0.007456566207110882
step: 90, loss: 0.0015602518105879426
step: 100, loss: 0.03288000449538231
step: 110, loss: 0.00857259426265955
step: 120, loss: 0.002538069849833846
step: 130, loss: 0.0798412561416626
step: 140, loss: 0.02367507293820381
step: 150, loss: 0.003096974454820156
step: 160, loss: 0.00258524133823812
step: 170, loss: 0.22690406441688538
step: 180, loss: 0.004460762720555067
step: 190, loss: 0.039063867181539536
step: 200, loss: 0.0343003012239933
step: 210, loss: 0.0038633786607533693
step: 220, loss: 0.024013807997107506
step: 230, loss: 0.09931506961584091
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 2: dev_f1=0.9841986455981941, f1=0.9830124575311437,
best_f1=0.9785794813979707
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005474865436553955
step: 10, loss: 0.1665625423192978
step: 20, loss: 0.005116167012602091
step: 30, loss: 0.016696076840162277
step: 40, loss: 0.0007469163974747062
step: 50, loss: 0.0005438752705231309
step: 60, loss: 0.0005491107003763318
step: 70, loss: 0.024006718769669533
step: 80, loss: 0.005018973257392645
step: 90, loss: 0.011449775658547878
step: 100, loss: 0.2105407863855362
step: 110, loss: 0.0845213383436203

```

```

step: 120, loss: 0.05516664311289787
step: 130, loss: 0.004396551754325628
step: 140, loss: 0.012924985960125923
step: 150, loss: 0.008714430965483189
step: 160, loss: 0.0005149453063495457
step: 170, loss: 0.00033915776293724775
step: 180, loss: 0.00017886958085000515
step: 190, loss: 0.049080390483140945
step: 200, loss: 0.19252605736255646
step: 210, loss: 0.017594560980796814
step: 220, loss: 0.014050985686480999
step: 230, loss: 0.0009583019418641925
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.9909706546275394, f1=0.9819819819819819,
best_f1=0.9819819819819819
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.011086470447480679
step: 10, loss: 0.009529327042400837
step: 20, loss: 0.0003064631891902536
step: 30, loss: 0.030500315129756927
step: 40, loss: 0.014466967433691025
step: 50, loss: 0.00042133466922678053
step: 60, loss: 0.001649864250794053
step: 70, loss: 0.0003691310703288764
step: 80, loss: 0.0012660472420975566
step: 90, loss: 0.01217185240238905
step: 100, loss: 0.004671911709010601
step: 110, loss: 0.00316448463127017
step: 120, loss: 0.0016575097106397152
step: 130, loss: 0.004971496760845184
step: 140, loss: 0.01152933482080698
step: 150, loss: 0.0018444611923769116
step: 160, loss: 0.00280209444463253
step: 170, loss: 0.0020278068259358406
step: 180, loss: 0.0008116188691928983
step: 190, loss: 0.002351585775613785
step: 200, loss: 0.001350598642602563
step: 210, loss: 0.032178983092308044
step: 220, loss: 0.007631823420524597
step: 230, loss: 0.0014943277928978205

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.988814317673378, f1=0.9821428571428571,
best_f1=0.9819819819819819
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006378627731464803
step: 10, loss: 0.0004710608336608857
step: 20, loss: 0.001517685828730464
step: 30, loss: 0.0410640686750412
step: 40, loss: 0.003184820292517543
step: 50, loss: 0.001403531525284052
step: 60, loss: 0.0004978899960406125
step: 70, loss: 0.00020443725225050002
step: 80, loss: 0.0005210074014030397
step: 90, loss: 0.00042983738239854574
step: 100, loss: 0.0014832951128482819
step: 110, loss: 0.0005004669073969126
step: 120, loss: 0.0010096057085320354
step: 130, loss: 0.0016373612452298403
step: 140, loss: 0.0013441054616123438
step: 150, loss: 0.00030338295619003475
step: 160, loss: 9.200807835441083e-05
step: 170, loss: 0.00030295111355371773
step: 180, loss: 0.0006351353949867189
step: 190, loss: 0.006948334630578756
step: 200, loss: 0.0003475573321338743
step: 210, loss: 0.003613697364926338
step: 220, loss: 0.00013354812108445913
step: 230, loss: 0.00010428536916151643
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.9932279909706545, f1=0.9886877828054299,
best_f1=0.9886877828054299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0033378670923411846
step: 10, loss: 0.004854223225265741
step: 20, loss: 0.0001525709667475894
step: 30, loss: 0.00032073090551421046

```



```

step: 40, loss: 0.0007863407372497022
step: 50, loss: 0.00016844173660501838
step: 60, loss: 0.0001577497459948063
step: 70, loss: 0.024346355348825455
step: 80, loss: 0.00010487436520634219
step: 90, loss: 0.00011150160571560264
step: 100, loss: 0.00016203121049329638
step: 110, loss: 0.00042868941091001034
step: 120, loss: 0.09303200989961624
step: 130, loss: 0.00017375141032971442
step: 140, loss: 0.001213973155245185
step: 150, loss: 0.00033635072759352624
step: 160, loss: 0.00017251809185836464
step: 170, loss: 0.000332443363731727
step: 180, loss: 0.002082634484395385
step: 190, loss: 0.0030933329835534096
step: 200, loss: 0.000358670688001439
step: 210, loss: 0.0003214096650481224
step: 220, loss: 0.00046912170364521444
step: 230, loss: 0.00024803398991934955
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.9886877828054299, f1=0.9829351535836178,
best_f1=0.9886877828054299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00029271558742038906
step: 10, loss: 0.0017346821259707212
step: 20, loss: 0.0008226933423429728
step: 30, loss: 0.0004010555276181549
step: 40, loss: 0.0006247934652492404
step: 50, loss: 0.00020487541041802615
step: 60, loss: 0.02659720741212368
step: 70, loss: 0.0009322637342847884
step: 80, loss: 0.0009874477982521057
step: 90, loss: 0.0005308687104843557
step: 100, loss: 0.0001489429414505139
step: 110, loss: 0.000112299392640125
step: 120, loss: 6.0126996686449274e-05
step: 130, loss: 0.0006515033892355859
step: 140, loss: 0.0001273597008548677
step: 150, loss: 0.0010162771213799715
step: 160, loss: 0.00011529375478858128
step: 170, loss: 0.0006310353055596352
step: 180, loss: 0.00012469627836253494
step: 190, loss: 0.0007030582055449486

```

```

step: 200, loss: 0.00023007579147815704
step: 210, loss: 0.030108382925391197
step: 220, loss: 0.0014713064301759005
step: 230, loss: 0.003012841334566474
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9887133182844244, f1=0.9875706214689265,
best_f1=0.9886877828054299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002889792376663536
step: 10, loss: 0.019174598157405853
step: 20, loss: 0.00033818528754636645
step: 30, loss: 0.00014233365072868764
step: 40, loss: 0.00013648314052261412
step: 50, loss: 0.0001964929251698777
step: 60, loss: 0.0003046246711164713
step: 70, loss: 0.00014102790737524629
step: 80, loss: 0.00018199595797341317
step: 90, loss: 0.00024303417012561113
step: 100, loss: 7.600900426041335e-05
step: 110, loss: 0.0003154359292238951
step: 120, loss: 0.001145069138146937
step: 130, loss: 4.9013604439096525e-05
step: 140, loss: 0.00018162188644055277
step: 150, loss: 4.8835026973392814e-05
step: 160, loss: 5.832176248077303e-05
step: 170, loss: 7.387882942566648e-05
step: 180, loss: 7.153573824325576e-05
step: 190, loss: 9.519905870547518e-05
step: 200, loss: 7.549781003035605e-05
step: 210, loss: 0.0003942936018574983
step: 220, loss: 6.616057362407446e-05
step: 230, loss: 0.00019295929814688861
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9899441340782122, f1=0.9866071428571428,
best_f1=0.9886877828054299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00015919622092042118
step: 10, loss: 4.470662315725349e-05
step: 20, loss: 0.0004000866028945893
step: 30, loss: 7.411149999825284e-05

```

```

step: 40, loss: 0.00016356896958313882
step: 50, loss: 7.665803423151374e-05
step: 60, loss: 0.000328255002386868
step: 70, loss: 3.75634990632534e-05
step: 80, loss: 0.02216995880007744
step: 90, loss: 0.00014565311721526086
step: 100, loss: 4.9918642616830766e-05
step: 110, loss: 0.00036370588350109756
step: 120, loss: 0.03676297515630722
step: 130, loss: 8.59969004523009e-05
step: 140, loss: 0.016718752682209015
step: 150, loss: 8.552028884878382e-05
step: 160, loss: 0.0007760411826893687
step: 170, loss: 0.00011339889169903472
step: 180, loss: 7.956975605338812e-05
step: 190, loss: 0.050117120146751404
step: 200, loss: 5.766655158367939e-05
step: 210, loss: 5.0569986342452466e-05
step: 220, loss: 0.00016699389379937202
step: 230, loss: 8.98573562153615e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.987709497206704, f1=0.987709497206704,
best_f1=0.9886877828054299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00011254589480813593
step: 10, loss: 0.07737503945827484
step: 20, loss: 0.02407778427004814
step: 30, loss: 9.117781155509874e-05
step: 40, loss: 0.012481140904128551
step: 50, loss: 6.558348832186311e-05
step: 60, loss: 5.411436359281652e-05
step: 70, loss: 0.007244954816997051
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 80, loss: 0.0010305955074727535
step: 90, loss: 0.0030657860916107893
step: 100, loss: 0.004244709853082895
step: 110, loss: 0.0022761384025216103
step: 120, loss: 0.0005553579539991915
step: 130, loss: 0.0005089387996122241
step: 140, loss: 5.657662404701114e-05
step: 150, loss: 6.920023588463664e-05
step: 160, loss: 0.037756942212581635
step: 170, loss: 4.5650394895346835e-05
step: 180, loss: 5.856850111740641e-05

```

```

step: 190, loss: 0.00017092069901991636
step: 200, loss: 0.00019477475143503398
step: 210, loss: 0.0005457366351038218
step: 220, loss: 0.00021019528503529727
step: 230, loss: 0.003059858689084649
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9886621315192743, f1=0.9875706214689265,
best_f1=0.9886877828054299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00020700800814665854
step: 10, loss: 7.793188706273213e-05
step: 20, loss: 4.8316742322640494e-05
step: 30, loss: 4.400830584927462e-05
step: 40, loss: 9.35910502448678e-05
step: 50, loss: 0.0004143654659856111
step: 60, loss: 0.011146625503897667
step: 70, loss: 4.735700713354163e-05
step: 80, loss: 0.00047222504508681595
step: 90, loss: 0.0006144965882413089
step: 100, loss: 7.553527393611148e-05
step: 110, loss: 6.892906094435602e-05
step: 120, loss: 9.211262658936903e-05
step: 130, loss: 0.0001361570757580921
step: 140, loss: 0.04084160178899765
step: 150, loss: 0.01627528853714466
step: 160, loss: 0.002112573478370905
step: 170, loss: 8.259124297183007e-05
step: 180, loss: 9.083235636353493e-05
step: 190, loss: 0.00023389408306684345
step: 200, loss: 0.012071794830262661
step: 210, loss: 0.0012896996922791004
step: 220, loss: 0.00010058098268928006
step: 230, loss: 4.082845407538116e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9899441340782122, f1=0.9855072463768116,
best_f1=0.9886877828054299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001282248122151941
step: 10, loss: 6.753917841706425e-05
step: 20, loss: 0.008150204084813595

```

```

step: 30, loss: 6.285265408223495e-05
step: 40, loss: 9.098363807424903e-05
step: 50, loss: 8.772189903538674e-05
step: 60, loss: 6.060715895728208e-05
step: 70, loss: 0.04915393516421318
step: 80, loss: 0.0004444087971933186
step: 90, loss: 9.724361007101834e-05
step: 100, loss: 7.955669570947066e-05
step: 110, loss: 6.496449350379407e-05
step: 120, loss: 0.04593160003423691
step: 130, loss: 7.2611597715877e-05
step: 140, loss: 0.00022764493769500405
step: 150, loss: 0.00011390849977033213
step: 160, loss: 8.338969200849533e-05
step: 170, loss: 0.00010792012471938506
step: 180, loss: 0.001477571320720017
step: 190, loss: 0.00034702118136920035
step: 200, loss: 0.00011890908353962004
step: 210, loss: 0.0215883981436491
step: 220, loss: 3.889813888235949e-05
step: 230, loss: 8.820120274322107e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9932279909706545, f1=0.9864864864864865,
best_f1=0.9886877828054299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.2898822913412e-05
step: 10, loss: 7.642269338248298e-05
step: 20, loss: 0.0005646975478157401
step: 30, loss: 0.00010018583998316899
step: 40, loss: 0.0004821563488803804
step: 50, loss: 0.00012600196350831538
step: 60, loss: 9.186148963635787e-05
step: 70, loss: 0.0006821214337833226
step: 80, loss: 9.15465789148584e-05
step: 90, loss: 9.560609032632783e-05
step: 100, loss: 7.045599340926856e-05
step: 110, loss: 0.00011073085624957457
step: 120, loss: 7.275417738128453e-05
step: 130, loss: 7.380857277894393e-05
step: 140, loss: 5.996409890940413e-05
step: 150, loss: 0.0031551579013466835
step: 160, loss: 0.00015641708159819245
step: 170, loss: 0.00036590074887499213
step: 180, loss: 8.753517613513395e-05

```

```

step: 190, loss: 4.886773967882618e-05
step: 200, loss: 7.429828110616654e-05
step: 210, loss: 0.00033970584627240896
step: 220, loss: 5.855323979631066e-05
step: 230, loss: 4.889874253422022e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9932279909706545, f1=0.9898305084745763,
best_f1=0.9886877828054299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.515727727673948e-05
step: 10, loss: 0.0002662611077539623
step: 20, loss: 6.61663943901658e-05
step: 30, loss: 7.947820995468646e-05
step: 40, loss: 3.141073102597147e-05
step: 50, loss: 6.109663081588224e-05
step: 60, loss: 8.577978587709367e-05
step: 70, loss: 4.7509507567156106e-05
step: 80, loss: 5.575195973506197e-05
step: 90, loss: 6.361518171615899e-05
step: 100, loss: 4.30128384323325e-05
step: 110, loss: 8.104691369226202e-05
step: 120, loss: 6.265858246479183e-05
step: 130, loss: 5.0203972932649776e-05
step: 140, loss: 0.014550680294632912
step: 150, loss: 5.4180389270186424e-05
step: 160, loss: 5.338379560271278e-05
step: 170, loss: 8.450964378425851e-05
step: 180, loss: 5.1212136895628646e-05
step: 190, loss: 0.00010355130507377908
step: 200, loss: 4.226519740768708e-05
step: 210, loss: 0.00013759649300482124
step: 220, loss: 0.015889201313257217
step: 230, loss: 0.00011089054169133306
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9932279909706545, f1=0.9887133182844244,
best_f1=0.9886877828054299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00011362327495589852
step: 10, loss: 3.6510064091999084e-05
step: 20, loss: 0.00011408528371248394

```

```

step: 30, loss: 0.0001227866450790316
step: 40, loss: 0.0046119000762701035
step: 50, loss: 4.562132016872056e-05
step: 60, loss: 0.00011994199303444475
step: 70, loss: 0.00010731482325354591
step: 80, loss: 0.0006276657804846764
step: 90, loss: 4.3337917304597795e-05
step: 100, loss: 0.0002904058783315122
step: 110, loss: 5.889611202292144e-05
step: 120, loss: 0.0325220450758934
step: 130, loss: 4.822657865588553e-05
step: 140, loss: 4.414615250425413e-05
step: 150, loss: 5.3603758715325966e-05
step: 160, loss: 0.008948828093707561
step: 170, loss: 5.597565905191004e-05
step: 180, loss: 0.00012034521932946518
step: 190, loss: 0.032982029020786285
step: 200, loss: 7.719115092186257e-05
step: 210, loss: 4.850421828450635e-05
step: 220, loss: 6.008490890963003e-05
step: 230, loss: 7.952060695970431e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 15: dev_f1=0.9932432432432432, f1=0.987598647125141,
best_f1=0.987598647125141
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type       : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32   : True
master_weights        : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type       : torch.float16
patch_torch_functions : False

```

```

keep_batchnorm_fp32      : True
master_weights           : True
loss_scale                : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2473it [00:07, 330.89it/s]
load_f1 = 0.9909502262443439
real_f1 = 0.9898305084745763
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:10, 402.73it/s]

```

3.1.3 DBLP-GoogleScholar - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
--task Structured/DBLP-GoogleScholar \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \
--dk product \
--summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-GoogleScholar.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \

```



```

--task Structured/DBLP-GoogleScholar \
--input_path input/input_small.jsonl \
--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:

UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order: `optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

```

"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)

```

```
step: 0, loss: 0.8869763612747192
```

```
step: 10, loss: 0.5445117354393005
```

```
step: 20, loss: 0.3654897212982178
```

step: 30, loss: 0.28241124749183655
step: 40, loss: 0.1825171262025833
step: 50, loss: 0.2613087296485901
step: 60, loss: 0.09233996272087097
step: 70, loss: 0.04840855300426483
step: 80, loss: 0.2221657931804657
step: 90, loss: 0.06281334906816483
step: 100, loss: 0.026978081092238426
step: 110, loss: 0.15597517788410187
step: 120, loss: 0.043944939970970154
step: 130, loss: 0.17502179741859436
step: 140, loss: 0.060710515826940536
step: 150, loss: 0.01810409128665924
step: 160, loss: 0.05039326101541519
step: 170, loss: 0.057667046785354614
step: 180, loss: 0.1576380878686905
step: 190, loss: 0.09542667120695114
step: 200, loss: 0.07851962745189667
step: 210, loss: 0.030520111322402954
step: 220, loss: 0.1333237886428833
step: 230, loss: 0.021206537261605263
step: 240, loss: 0.046588774770498276
step: 250, loss: 0.169707253575325
step: 260, loss: 0.06409629434347153
step: 270, loss: 0.013066999614238739
step: 280, loss: 0.009332822635769844
step: 290, loss: 0.006493670400232077
step: 300, loss: 0.018841836601495743
step: 310, loss: 0.08028089255094528
step: 320, loss: 0.01759457215666771
step: 330, loss: 0.14940796792507172
step: 340, loss: 0.03355477750301361
step: 350, loss: 0.040813468396663666
step: 360, loss: 0.11256904155015945
step: 370, loss: 0.05615146830677986
step: 380, loss: 0.023085888475179672
step: 390, loss: 0.024497143924236298
step: 400, loss: 0.027322182431817055
step: 410, loss: 0.1157914474606514
step: 420, loss: 0.03603677824139595
step: 430, loss: 0.20446842908859253
step: 440, loss: 0.06591338664293289
step: 450, loss: 0.09264534711837769
step: 460, loss: 0.053829293698072433
step: 470, loss: 0.025023162364959717
step: 480, loss: 0.01789727620780468
step: 490, loss: 0.04434813931584358
step: 500, loss: 0.14354896545410156

```

step: 510, loss: 0.03142286837100983
step: 520, loss: 0.054349515587091446
step: 530, loss: 0.05949327349662781
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9496470588235294, f1=0.9427230046948356,
best_f1=0.9427230046948356
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01130965817719698
step: 10, loss: 0.012799397110939026
step: 20, loss: 0.008125502616167068
step: 30, loss: 0.018023919314146042
step: 40, loss: 0.1683611124753952
step: 50, loss: 0.014152472838759422
step: 60, loss: 0.06475843489170074
step: 70, loss: 0.09985315799713135
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 80, loss: 0.047671061009168625
step: 90, loss: 0.06993945688009262
step: 100, loss: 0.18219415843486786
step: 110, loss: 0.09387605637311935
step: 120, loss: 0.029788799583911896
step: 130, loss: 0.06990395486354828
step: 140, loss: 0.025266356766223907
step: 150, loss: 0.007071506232023239
step: 160, loss: 0.008809199556708336
step: 170, loss: 0.09643186628818512
step: 180, loss: 0.10100453346967697
step: 190, loss: 0.0797649472951889
step: 200, loss: 0.0400875099003315
step: 210, loss: 0.0271461084485054
step: 220, loss: 0.01722288504242897
step: 230, loss: 0.04039822518825531
step: 240, loss: 0.016338268294930458
step: 250, loss: 0.07806295901536942
step: 260, loss: 0.008581814356148243
step: 270, loss: 0.06235915422439575
step: 280, loss: 0.13072943687438965
step: 290, loss: 0.03232980892062187
step: 300, loss: 0.04102287441492081
step: 310, loss: 0.007755434140563011

```

```

step: 320, loss: 0.08983499556779861
step: 330, loss: 0.015133578330278397
step: 340, loss: 0.0038472090382128954
step: 350, loss: 0.013041088357567787
step: 360, loss: 0.06968312710523605
step: 370, loss: 0.12235748767852783
step: 380, loss: 0.08678212761878967
step: 390, loss: 0.03807881101965904
step: 400, loss: 0.019850391894578934
step: 410, loss: 0.0024817625526338816
step: 420, loss: 0.002637189580127597
step: 430, loss: 0.036607179790735245
step: 440, loss: 0.008410533890128136
step: 450, loss: 0.05344812572002411
step: 460, loss: 0.2093597799539566
step: 470, loss: 0.08520045876502991
step: 480, loss: 0.015685485675930977
step: 490, loss: 0.18380653858184814
step: 500, loss: 0.004615431651473045
step: 510, loss: 0.026933444663882256
step: 520, loss: 0.122479647397995
step: 530, loss: 0.11714759469032288
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 2: dev_f1=0.9450139794967382, f1=0.9433611884865365,
best_f1=0.9427230046948356
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.031009363010525703
step: 10, loss: 0.004331938456743956
step: 20, loss: 0.006785261910408735
step: 30, loss: 0.0019937637262046337
step: 40, loss: 0.003264077939093113
step: 50, loss: 0.0046156588941812515
step: 60, loss: 0.017115222290158272
step: 70, loss: 0.041850972920656204
step: 80, loss: 0.06445471197366714
step: 90, loss: 0.024249952286481857
step: 100, loss: 0.002195958746597171
step: 110, loss: 0.04507585987448692
step: 120, loss: 0.0453094057738781
step: 130, loss: 0.033112186938524246
step: 140, loss: 0.000692669244017452
step: 150, loss: 0.03206346929073334
step: 160, loss: 0.010882499627768993
step: 170, loss: 0.0013497250620275736

```

```

step: 180, loss: 0.059138379991054535
step: 190, loss: 0.0017280681058764458
step: 200, loss: 0.005709419026970863
step: 210, loss: 0.04407927766442299
step: 220, loss: 0.03260234743356705
step: 230, loss: 0.006495218724012375
step: 240, loss: 0.0006852729129604995
step: 250, loss: 0.02303069643676281
step: 260, loss: 0.006112953182309866
step: 270, loss: 0.02771061472594738
step: 280, loss: 0.046902257949113846
step: 290, loss: 0.05109826847910881
step: 300, loss: 0.09957096725702286
step: 310, loss: 0.01571439579129219
step: 320, loss: 0.02949623204767704
step: 330, loss: 0.009965328499674797
step: 340, loss: 0.045186806470155716
step: 350, loss: 0.00368228810839355
step: 360, loss: 0.010843567550182343
step: 370, loss: 0.019454987719655037
step: 380, loss: 0.09403113275766373
step: 390, loss: 0.017987284809350967
step: 400, loss: 0.024159489199519157
step: 410, loss: 0.0009065442136488855
step: 420, loss: 0.03183763101696968
step: 430, loss: 0.019866013899445534
step: 440, loss: 0.1696920394897461
step: 450, loss: 0.005384575575590134
step: 460, loss: 0.003302905475720763
step: 470, loss: 0.0017473582411184907
step: 480, loss: 0.07279743999242783
step: 490, loss: 0.05289836972951889
step: 500, loss: 0.03569312393665314
step: 510, loss: 0.014197679236531258
step: 520, loss: 0.03887277841567993
step: 530, loss: 0.10016992688179016
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.944874715261959, f1=0.9418181818181819,
best_f1=0.9427230046948356
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0037828192580491304
step: 10, loss: 0.013288665562868118
step: 20, loss: 0.0014959872933104634
step: 30, loss: 9.359502291772515e-05

```

step: 40, loss: 0.010546710342168808
step: 50, loss: 0.0066011399030685425
step: 60, loss: 0.000850232841912657
step: 70, loss: 0.00030324343242682517
step: 80, loss: 0.0022870854008942842
step: 90, loss: 0.039011821150779724
step: 100, loss: 0.08926733583211899
step: 110, loss: 0.0005965972668491304
step: 120, loss: 0.04007413983345032
step: 130, loss: 0.00098600413184613
step: 140, loss: 0.0005443515838123858
step: 150, loss: 0.0018760313978418708
step: 160, loss: 0.00023249807418324053
step: 170, loss: 0.0005840658559463918
step: 180, loss: 0.030576353892683983
step: 190, loss: 0.0006871507503092289
step: 200, loss: 0.1489664763212204
step: 210, loss: 0.10367445647716522
step: 220, loss: 0.011843690648674965
step: 230, loss: 0.06390024721622467
step: 240, loss: 0.08586553484201431
step: 250, loss: 0.012380487285554409
step: 260, loss: 0.006686362903565168
step: 270, loss: 0.014592309482395649
step: 280, loss: 0.009843415580689907
step: 290, loss: 0.0004727535124402493
step: 300, loss: 0.0010571200400590897
step: 310, loss: 0.0022427840158343315
step: 320, loss: 0.06557005643844604
step: 330, loss: 0.005317377857863903
step: 340, loss: 0.007552947849035263
step: 350, loss: 0.0003223264357075095
step: 360, loss: 0.008637445978820324
step: 370, loss: 0.006778533570468426
step: 380, loss: 0.010831153951585293
step: 390, loss: 0.003955053165555
step: 400, loss: 0.030130859464406967
step: 410, loss: 0.02424004301428795
step: 420, loss: 0.0014100144617259502
step: 430, loss: 0.04930174723267555
step: 440, loss: 5.4109848861116916e-05
step: 450, loss: 0.0361793227493763
step: 460, loss: 0.00026063912082463503
step: 470, loss: 0.0005175896803848445
step: 480, loss: 0.016875291243195534
step: 490, loss: 0.0019773163367062807
step: 500, loss: 0.006620489992201328
step: 510, loss: 0.0013650452019646764

```

step: 520, loss: 0.006266618613153696
step: 530, loss: 0.0005320325726643205
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.9436749769159741, f1=0.944649446494465,
best_f1=0.9427230046948356
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0018726503476500511
step: 10, loss: 0.17516587674617767
step: 20, loss: 0.12051918357610703
step: 30, loss: 0.008046326227486134
step: 40, loss: 0.009272610768675804
step: 50, loss: 0.0041399323381483555
step: 60, loss: 0.00893489085137844
step: 70, loss: 0.011745831929147243
step: 80, loss: 0.004797056782990694
step: 90, loss: 0.011287814937531948
step: 100, loss: 0.00022790423827245831
step: 110, loss: 0.0018014064989984035
step: 120, loss: 0.006867780815809965
step: 130, loss: 0.001436376478523016
step: 140, loss: 0.004322669003158808
step: 150, loss: 0.0015112929977476597
step: 160, loss: 0.00012842653086408973
step: 170, loss: 0.0006949318922124803
step: 180, loss: 0.004470109939575195
step: 190, loss: 0.00023882053210400045
step: 200, loss: 0.00047346888459287584
step: 210, loss: 0.00018577053560875356
step: 220, loss: 0.0020630848594009876
step: 230, loss: 0.05916757509112358
step: 240, loss: 0.000959194905590266
step: 250, loss: 0.010021707974374294
step: 260, loss: 0.00017338884936179966
step: 270, loss: 0.009612619876861572
step: 280, loss: 0.004594855941832066
step: 290, loss: 0.012890320271253586
step: 300, loss: 0.00015743100084364414
step: 310, loss: 9.707050048746169e-05
step: 320, loss: 0.00011103373253718019
step: 330, loss: 0.0014704653294757009
step: 340, loss: 0.037034474313259125
step: 350, loss: 0.0003180786152370274
step: 360, loss: 0.006996353156864643
step: 370, loss: 0.008036733604967594

```

```
step: 380, loss: 0.0002732613938860595
step: 390, loss: 0.0005445662536658347
step: 400, loss: 0.0012038189452141523
step: 410, loss: 0.00038296563434414566
step: 420, loss: 0.03626095876097679
step: 430, loss: 0.056319914758205414
step: 440, loss: 0.005378626752644777
step: 450, loss: 0.00021111808018758893
step: 460, loss: 0.00032387266401201487
step: 470, loss: 0.03243101015686989
step: 480, loss: 0.0008522423449903727
step: 490, loss: 0.004868841264396906
step: 500, loss: 0.0014859420480206609
step: 510, loss: 0.014597763307392597
step: 520, loss: 0.00027074164245277643
step: 530, loss: 0.004587805364280939
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.9448373408769447, f1=0.9449626865671642,
best_f1=0.9427230046948356
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0010873363353312016
step: 10, loss: 0.003498418489471078
step: 20, loss: 0.008103826083242893
step: 30, loss: 0.00045541333383880556
step: 40, loss: 0.00010534566536080092
step: 50, loss: 0.0023524572607129812
step: 60, loss: 0.0001502337836427614
step: 70, loss: 0.00764368986710906
step: 80, loss: 0.0005431046593002975
step: 90, loss: 0.00011321562487864867
step: 100, loss: 0.0014458289369940758
step: 110, loss: 0.001643608440645039
step: 120, loss: 0.0005239429301582277
step: 130, loss: 0.008822184056043625
step: 140, loss: 6.218505586730316e-05
step: 150, loss: 8.779024210525677e-05
step: 160, loss: 0.0003327786398585886
step: 170, loss: 0.0001755519479047507
step: 180, loss: 0.00013487046817317605
step: 190, loss: 0.003065031487494707
step: 200, loss: 0.0025012234691530466
step: 210, loss: 0.01752694509923458
step: 220, loss: 0.0012870461214333773
step: 230, loss: 0.000177496942342259
```



```

step: 240, loss: 0.06426499783992767
step: 250, loss: 7.503574306610972e-05
step: 260, loss: 0.0028550440911203623
step: 270, loss: 0.002924219472333789
step: 280, loss: 0.000544803449884057
step: 290, loss: 0.0006589205586351454
step: 300, loss: 0.0010473649017512798
step: 310, loss: 0.000117857736768201
step: 320, loss: 0.00038116826908662915
step: 330, loss: 0.03916572779417038
step: 340, loss: 0.000537087325938046
step: 350, loss: 0.0051406691782176495
step: 360, loss: 0.0018444242887198925
step: 370, loss: 7.922016811789945e-05
step: 380, loss: 0.0006966334185563028
step: 390, loss: 0.012726884335279465
step: 400, loss: 0.004947138484567404
step: 410, loss: 5.304247315507382e-05
step: 420, loss: 0.0007302496815100312
step: 430, loss: 0.0005691641708835959
step: 440, loss: 8.293290011351928e-05
step: 450, loss: 0.017075497657060623
step: 460, loss: 0.00016650924226269126
step: 470, loss: 0.0003277773503214121
step: 480, loss: 0.0007809267262928188
step: 490, loss: 0.00016815579147078097
step: 500, loss: 7.02500983607024e-05
step: 510, loss: 0.00010607689910102636
step: 520, loss: 0.0001819870522012934
step: 530, loss: 0.00201828102581203
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.9435559736594544, f1=0.9424964936886395,
best_f1=0.9427230046948356
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0018856627866625786
step: 10, loss: 0.021209757775068283
step: 20, loss: 0.00013146529090590775
step: 30, loss: 0.06141705811023712
step: 40, loss: 0.027377383783459663
step: 50, loss: 0.004195543471723795
step: 60, loss: 0.0012572318082675338
step: 70, loss: 0.0005937029491178691
step: 80, loss: 8.990331116365269e-05
step: 90, loss: 0.0014709531096741557

```

```
step: 100, loss: 0.0001249753695446998
step: 110, loss: 0.0024312706664204597
step: 120, loss: 0.00023311674885917455
step: 130, loss: 0.0024826196022331715
step: 140, loss: 0.007588403765112162
step: 150, loss: 0.00038790900725871325
step: 160, loss: 0.00013602034596260637
step: 170, loss: 0.005354125984013081
step: 180, loss: 0.00012114347191527486
step: 190, loss: 0.001997669693082571
step: 200, loss: 0.003353042993694544
step: 210, loss: 0.0010555833578109741
step: 220, loss: 0.00037651159800589085
step: 230, loss: 0.003175335004925728
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 240, loss: 0.0005059405812062323
step: 250, loss: 0.0008919791434891522
step: 260, loss: 0.00015972043911460787
step: 270, loss: 0.001005577272735536
step: 280, loss: 0.00042590219527482986
step: 290, loss: 0.005454955156892538
step: 300, loss: 0.007465839851647615
step: 310, loss: 0.00015005512977950275
step: 320, loss: 0.00047082791570574045
step: 330, loss: 0.0182887502014637
step: 340, loss: 4.320712105254643e-05
step: 350, loss: 0.0019504155497998
step: 360, loss: 0.0039302813820540905
step: 370, loss: 3.680440931930207e-05
step: 380, loss: 0.006593676749616861
step: 390, loss: 0.020730705931782722
step: 400, loss: 0.004067016765475273
step: 410, loss: 6.598564505111426e-05
step: 420, loss: 0.00016865538782440126
step: 430, loss: 0.00880133081227541
step: 440, loss: 0.0064765592105686665
step: 450, loss: 0.0003189309500157833
step: 460, loss: 0.0003110092948190868
step: 470, loss: 0.05915798246860504
step: 480, loss: 0.0014156585093587637
step: 490, loss: 8.68832430569455e-05
step: 500, loss: 0.002992664696648717
step: 510, loss: 0.0019024154171347618
step: 520, loss: 0.00010903801012318581
step: 530, loss: 0.0008584044990129769
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
epoch 7: dev_f1=0.9486940298507462, f1=0.9480519480519481,  
best_f1=0.9427230046948356  
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:  
An input tensor was not cuda.  
    warnings.warn("An input tensor was not cuda.")  
step: 0, loss: 0.007213684264570475  
step: 10, loss: 0.0001651119237067178  
step: 20, loss: 8.822032395983115e-05  
step: 30, loss: 0.00011454836203483865  
step: 40, loss: 5.543517181649804e-05  
step: 50, loss: 3.4472657716833055e-05  
step: 60, loss: 0.0020516382064670324  
step: 70, loss: 8.194647671189159e-05  
step: 80, loss: 0.001368801691569388  
step: 90, loss: 0.0002266008232254535  
step: 100, loss: 0.0016899786423891783  
step: 110, loss: 4.201752017252147e-05  
step: 120, loss: 0.0002296984166605398  
step: 130, loss: 4.673911462305114e-05  
step: 140, loss: 4.4566990254679695e-05  
step: 150, loss: 0.00040699689998291433  
step: 160, loss: 0.15991224348545074  
step: 170, loss: 0.003459957428276539  
step: 180, loss: 0.004580297041684389  
step: 190, loss: 0.002606919500976801  
step: 200, loss: 7.752818055450916e-05  
step: 210, loss: 5.4676507716067135e-05  
step: 220, loss: 0.00033747454290278256  
step: 230, loss: 0.00260428199544549  
step: 240, loss: 6.2934348534327e-05  
step: 250, loss: 0.0018615782028064132  
step: 260, loss: 3.6569930671248585e-05  
step: 270, loss: 0.06619870662689209  
step: 280, loss: 3.129902688669972e-05  
step: 290, loss: 0.0010437496239319444  
step: 300, loss: 0.0014693026896566153  
step: 310, loss: 0.00043168061529286206  
step: 320, loss: 5.176262857276015e-05  
step: 330, loss: 5.99073464400135e-05  
step: 340, loss: 0.001924841315485537  
step: 350, loss: 7.235190423671156e-05  
step: 360, loss: 0.004414767492562532  
step: 370, loss: 0.20027956366539001  
step: 380, loss: 0.0015174539294093847  
step: 390, loss: 0.12369890511035919  
step: 400, loss: 0.03144681081175804  
step: 410, loss: 0.0009194801095873117  
step: 420, loss: 0.0012956527061760426
```

```

step: 430, loss: 0.003953469917178154
step: 440, loss: 0.004478679038584232
step: 450, loss: 0.0006459050928242505
step: 460, loss: 0.0041092680767178535
step: 470, loss: 0.0024757892824709415
step: 480, loss: 0.018212534487247467
step: 490, loss: 0.00012762733967974782
step: 500, loss: 0.037623658776283264
step: 510, loss: 0.014935044571757317
step: 520, loss: 0.0003912326064892113
step: 530, loss: 0.00090879196068272
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9441391941391942, f1=0.9441163107678328,
best_f1=0.9427230046948356
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003620445786509663
step: 10, loss: 0.0001560647797305137
step: 20, loss: 0.00019330313079990447
step: 30, loss: 0.004983728751540184
step: 40, loss: 4.5681954361498356e-05
step: 50, loss: 0.0028271919582039118
step: 60, loss: 0.0008217219146899879
step: 70, loss: 0.000376153300749138
step: 80, loss: 0.0013956077164039016
step: 90, loss: 4.3903612095164135e-05
step: 100, loss: 0.00014542473945766687
step: 110, loss: 5.557075564865954e-05
step: 120, loss: 0.024186454713344574
step: 130, loss: 0.00018642925715539604
step: 140, loss: 0.006970972754061222
step: 150, loss: 3.434616883168928e-05
step: 160, loss: 9.161562047665939e-05
step: 170, loss: 2.2965970856603235e-05
step: 180, loss: 5.633952605421655e-05
step: 190, loss: 0.00035414600279182196
step: 200, loss: 6.956681318115443e-05
step: 210, loss: 2.1323166947695427e-05
step: 220, loss: 0.004480571020394564
step: 230, loss: 0.00012602549395523965
step: 240, loss: 0.0015880069695413113
step: 250, loss: 0.0002161348966183141
step: 260, loss: 0.0005801641964353621
step: 270, loss: 3.638098496594466e-05
step: 280, loss: 0.029890138655900955

```

```

step: 290, loss: 4.626579539035447e-05
step: 300, loss: 0.0007561617530882359
step: 310, loss: 3.2462772651342675e-05
step: 320, loss: 0.02252253144979477
step: 330, loss: 0.106624536216259
step: 340, loss: 0.00015275159967131913
step: 350, loss: 9.997146116802469e-05
step: 360, loss: 0.0007150550372898579
step: 370, loss: 7.020822522463277e-05
step: 380, loss: 3.2408308470621705e-05
step: 390, loss: 3.948783705709502e-05
step: 400, loss: 0.00015203090151771903
step: 410, loss: 6.39821300865151e-05
step: 420, loss: 0.000583988381549716
step: 430, loss: 0.0030411924235522747
step: 440, loss: 0.000547543284483254
step: 450, loss: 4.146359788137488e-05
step: 460, loss: 3.3845582947833464e-05
step: 470, loss: 0.0006290934979915619
step: 480, loss: 0.0007368357037194073
step: 490, loss: 0.0024813213385641575
step: 500, loss: 0.0010959123028442264
step: 510, loss: 0.0007817756268195808
step: 520, loss: 0.0013249260373413563
step: 530, loss: 6.918835424585268e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9492787342950209, f1=0.9529085872576176,
best_f1=0.9427230046948356
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.021484103053808212
step: 10, loss: 0.0004341202147770673
step: 20, loss: 0.0003715330676641315
step: 30, loss: 3.566024315659888e-05
step: 40, loss: 7.780222222208977e-05
step: 50, loss: 6.91102395649068e-05
step: 60, loss: 5.769867493654601e-05
step: 70, loss: 4.31717635365203e-05
step: 80, loss: 2.1524294425034896e-05
step: 90, loss: 5.7289904361823574e-05
step: 100, loss: 0.0004583230766002089
step: 110, loss: 0.00086630490841344
step: 120, loss: 5.9179314121138304e-05
step: 130, loss: 2.668728120625019e-05
step: 140, loss: 0.011412303894758224

```

```
step: 150, loss: 0.001354794716462493
step: 160, loss: 0.07376469671726227
step: 170, loss: 0.0007093635504133999
step: 180, loss: 0.0028884599450975657
step: 190, loss: 5.526398672373034e-05
step: 200, loss: 0.01427686121314764
step: 210, loss: 0.0018532436806708574
step: 220, loss: 8.903567504603416e-05
step: 230, loss: 4.714237729785964e-05
step: 240, loss: 5.261931073619053e-05
step: 250, loss: 0.004060542210936546
step: 260, loss: 0.00038104443228803575
step: 270, loss: 6.587077950825915e-05
step: 280, loss: 0.0002427537547191605
step: 290, loss: 0.00010378183651482686
step: 300, loss: 0.00012813603098038584
step: 310, loss: 1.4968068171583582e-05
step: 320, loss: 0.0005444011185318232
step: 330, loss: 4.7290224756579846e-05
step: 340, loss: 7.554060721304268e-05
step: 350, loss: 0.0001244092418346554
step: 360, loss: 4.16096045228187e-05
step: 370, loss: 0.0024174952413886786
step: 380, loss: 0.0022571254521608353
step: 390, loss: 0.0008901843102648854
step: 400, loss: 0.00010094205208588392
step: 410, loss: 3.1923871574690565e-05
step: 420, loss: 2.7689075068337843e-05
step: 430, loss: 0.0001606350124347955
step: 440, loss: 0.00420331722125411
step: 450, loss: 0.0026729281526058912
step: 460, loss: 0.005290233995765448
step: 470, loss: 0.0014241774333640933
step: 480, loss: 0.08508144319057465
step: 490, loss: 0.0064236498437821865
step: 500, loss: 7.067016122164205e-05
step: 510, loss: 0.0020840433426201344
step: 520, loss: 0.0009792884811758995
step: 530, loss: 0.0010859450558200479
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9469732519943688, f1=0.9428303655107779,
best_f1=0.9427230046948356
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00025677154189907014
```

step: 10, loss: 0.00016104466340038925
step: 20, loss: 0.00110557209700346
step: 30, loss: 8.560059359297156e-05
step: 40, loss: 0.0014919271925464272
step: 50, loss: 0.032701876014471054
step: 60, loss: 0.08451119065284729
step: 70, loss: 9.481832239544019e-05
step: 80, loss: 0.0007214472279883921
step: 90, loss: 0.005992427002638578
step: 100, loss: 0.0009019735734909773
step: 110, loss: 0.0007704380550421774
step: 120, loss: 0.005997390486299992
step: 130, loss: 0.00017925810243468732
step: 140, loss: 0.04185497760772705
step: 150, loss: 2.9845507015124895e-05
step: 160, loss: 0.024691279977560043
step: 170, loss: 3.3667052775854245e-05
step: 180, loss: 0.00012187119864393026
step: 190, loss: 3.177893813699484e-05
step: 200, loss: 5.095350570627488e-05
step: 210, loss: 8.178248390322551e-05
step: 220, loss: 0.005540828220546246
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 230, loss: 8.804591197986156e-05
step: 240, loss: 8.255537977674976e-05
step: 250, loss: 4.1339517338201404e-05
step: 260, loss: 3.566369559848681e-05
step: 270, loss: 0.13454978168010712
step: 280, loss: 0.00011829817231046036
step: 290, loss: 0.0004402559425216168
step: 300, loss: 0.00013229945034254342
step: 310, loss: 1.9952047296101227e-05
step: 320, loss: 9.899672295432538e-05
step: 330, loss: 2.5345736503368244e-05
step: 340, loss: 0.00019195053027942777
step: 350, loss: 0.0006677326746284962
step: 360, loss: 1.7385411410941742e-05
step: 370, loss: 9.466190385865048e-05
step: 380, loss: 3.802934952545911e-05
step: 390, loss: 3.2888852729229257e-05
step: 400, loss: 5.489488466992043e-05
step: 410, loss: 2.875651807698887e-05
step: 420, loss: 4.385926513350569e-05
step: 430, loss: 0.0004627702583093196
step: 440, loss: 5.002175385016017e-05
step: 450, loss: 8.125607564579695e-05
step: 460, loss: 0.00022882106713950634
step: 470, loss: 4.753369648824446e-05

```

step: 480, loss: 0.00011066473962273449
step: 490, loss: 0.04523243010044098
step: 500, loss: 0.00041259080171585083
step: 510, loss: 1.6122816305141896e-05
step: 520, loss: 0.06600247323513031
step: 530, loss: 7.467361865565181e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9492957746478873, f1=0.9487895716945998,
best_f1=0.9427230046948356
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.223659198032692e-05
step: 10, loss: 3.007554005307611e-05
step: 20, loss: 9.051604138221592e-05
step: 30, loss: 0.00016070737910922617
step: 40, loss: 0.002643959829583764
step: 50, loss: 0.0003464510664343834
step: 60, loss: 7.015600567683578e-05
step: 70, loss: 0.00012119553866796196
step: 80, loss: 4.470000567380339e-05
step: 90, loss: 0.0036032055504620075
step: 100, loss: 2.747623329923954e-05
step: 110, loss: 6.623766239499673e-05
step: 120, loss: 2.8757936888723634e-05
step: 130, loss: 2.724923615460284e-05
step: 140, loss: 0.00020841405785176903
step: 150, loss: 0.007049825042486191
step: 160, loss: 4.8059806431410834e-05
step: 170, loss: 0.009795929305255413
step: 180, loss: 0.0005026712315157056
step: 190, loss: 0.04652766138315201
step: 200, loss: 6.591510464204475e-05
step: 210, loss: 0.00037854135734960437
step: 220, loss: 7.875082519603893e-05
step: 230, loss: 4.057227124576457e-05
step: 240, loss: 0.004070614464581013
step: 250, loss: 3.38139507221058e-05
step: 260, loss: 0.07394812256097794
step: 270, loss: 2.3476330170524307e-05
step: 280, loss: 4.331864329287782e-05
step: 290, loss: 4.5813314500264823e-05
step: 300, loss: 1.5664616512367502e-05
step: 310, loss: 2.160988333343994e-05
step: 320, loss: 1.934869578690268e-05
step: 330, loss: 2.6210453142994083e-05

```



```

step: 340, loss: 3.174848097842187e-05
step: 350, loss: 2.140127617167309e-05
step: 360, loss: 0.0003054890548810363
step: 370, loss: 0.0035136807709932327
step: 380, loss: 6.302937254076824e-05
step: 390, loss: 7.443727372447029e-05
step: 400, loss: 0.0006783455028198659
step: 410, loss: 0.00047518228529952466
step: 420, loss: 0.012870828621089458
step: 430, loss: 2.6303152480977587e-05
step: 440, loss: 2.483195930835791e-05
step: 450, loss: 0.002842494985088706
step: 460, loss: 2.3449842046829872e-05
step: 470, loss: 3.2538613595534116e-05
step: 480, loss: 0.04876786470413208
step: 490, loss: 0.00020407923148013651
step: 500, loss: 3.271264358772896e-05
step: 510, loss: 0.02354479767382145
step: 520, loss: 0.000135340858832933
step: 530, loss: 0.00023566227173432708
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.947764705882353, f1=0.9486461251167133,
best_f1=0.9427230046948356
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0023161580320447683
step: 10, loss: 2.0496026991168037e-05
step: 20, loss: 4.3340620322851464e-05
step: 30, loss: 0.00018720982188824564
step: 40, loss: 1.3407147889665794e-05
step: 50, loss: 0.01527189277112484
step: 60, loss: 1.951266312971711e-05
step: 70, loss: 1.6495294403284788e-05
step: 80, loss: 0.00013130529259797186
step: 90, loss: 1.5012695257610176e-05
step: 100, loss: 3.248152643209323e-05
step: 110, loss: 0.001983099151402712
step: 120, loss: 1.6640580724924803e-05
step: 130, loss: 0.0007235483499243855
step: 140, loss: 1.6428261005785316e-05
step: 150, loss: 0.00040648659341968596
step: 160, loss: 0.000403668760554865
step: 170, loss: 1.5318139048758894e-05
step: 180, loss: 2.6851035727304406e-05
step: 190, loss: 0.0008576602558605373

```

```

step: 200, loss: 0.00023826713731978089
step: 210, loss: 3.5661552828969434e-05
step: 220, loss: 0.054317303001880646
step: 230, loss: 1.910279388539493e-05
step: 240, loss: 0.0012938507134094834
step: 250, loss: 2.4470808057230897e-05
step: 260, loss: 3.988617754657753e-05
step: 270, loss: 0.0030435589142143726
step: 280, loss: 0.001167974667623639
step: 290, loss: 2.9197770345490426e-05
step: 300, loss: 0.000290411408059299
step: 310, loss: 2.001539724005852e-05
step: 320, loss: 0.00024344539269804955
step: 330, loss: 0.00013416854199022055
step: 340, loss: 2.4362661861232482e-05
step: 350, loss: 0.006995284929871559
step: 360, loss: 0.015887927263975143
step: 370, loss: 2.024646892095916e-05
step: 380, loss: 0.00043793718214146793
step: 390, loss: 1.9869910829584114e-05
step: 400, loss: 2.3397918994305655e-05
step: 410, loss: 2.0763731299666688e-05
step: 420, loss: 2.5628354705986567e-05
step: 430, loss: 0.00010861392365768552
step: 440, loss: 7.471840945072472e-05
step: 450, loss: 2.7774873160524294e-05
step: 460, loss: 5.7426852436037734e-05
step: 470, loss: 0.003644856158643961
step: 480, loss: 2.8157237466075458e-05
step: 490, loss: 0.0020832556765526533
step: 500, loss: 2.7052241421188228e-05
step: 510, loss: 0.0037878507282584906
step: 520, loss: 2.1360348910093307e-05
step: 530, loss: 2.1468189515871927e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 13: dev_f1=0.9512195121951219, f1=0.9510032664489033,
best_f1=0.9510032664489033
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 2.7521657102624886e-05
step: 10, loss: 1.760524173732847e-05

```

step: 20, loss: 2.967076397908386e-05
step: 30, loss: 2.8842914616689086e-05
step: 40, loss: 1.858882387750782e-05
step: 50, loss: 3.040394039999228e-05
step: 60, loss: 0.0001334106782451272
step: 70, loss: 2.918623067671433e-05
step: 80, loss: 1.954232357093133e-05
step: 90, loss: 0.00031182379461824894
step: 100, loss: 0.0008997328695841134
step: 110, loss: 0.0003218559722881764
step: 120, loss: 7.317709969356656e-05
step: 130, loss: 2.893947748816572e-05
step: 140, loss: 2.333081647520885e-05
step: 150, loss: 2.2857755539007485e-05
step: 160, loss: 0.011607762426137924
step: 170, loss: 2.2097699911682867e-05
step: 180, loss: 6.156910967547446e-05
step: 190, loss: 2.4813285563141108e-05
step: 200, loss: 1.7206821212312207e-05
step: 210, loss: 2.1729150830651633e-05
step: 220, loss: 1.758305006660521e-05
step: 230, loss: 3.176380414515734e-05
step: 240, loss: 2.1736588678322732e-05
step: 250, loss: 0.0003522606857586652
step: 260, loss: 0.04529057443141937
step: 270, loss: 9.059010335477069e-05
step: 280, loss: 1.909918137243949e-05
step: 290, loss: 3.194896271452308e-05
step: 300, loss: 0.004319115541875362
step: 310, loss: 4.460018681129441e-05
step: 320, loss: 3.053769614780322e-05
step: 330, loss: 4.0955990698421374e-05
step: 340, loss: 0.00121911836322397
step: 350, loss: 0.000688960135448724
step: 360, loss: 2.6783849534695037e-05
step: 370, loss: 2.492127168807201e-05
step: 380, loss: 0.00026483426336199045
step: 390, loss: 0.0001619593967916444
step: 400, loss: 2.6090865503647365e-05
step: 410, loss: 1.866710590547882e-05
step: 420, loss: 3.393330916878767e-05
step: 430, loss: 3.071017636102624e-05
step: 440, loss: 3.511260001687333e-05
step: 450, loss: 0.0027891495265066624
step: 460, loss: 4.303721652831882e-05
step: 470, loss: 2.299943662364967e-05
step: 480, loss: 2.315894926141482e-05
step: 490, loss: 2.9554472348536365e-05

```

step: 500, loss: 2.270125514769461e-05
step: 510, loss: 3.339532850077376e-05
step: 520, loss: 4.8867135774344206e-05
step: 530, loss: 2.7036267056246288e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 14: dev_f1=0.9517177344475395, f1=0.9539594843462248,
best_f1=0.9539594843462248
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 2.53274010901805e-05
step: 10, loss: 1.7311051124124788e-05
step: 20, loss: 2.5408953661099076e-05
step: 30, loss: 8.971260103862733e-05
step: 40, loss: 0.00023656779376324266
step: 50, loss: 4.106792403035797e-05
step: 60, loss: 0.0009897354757413268
step: 70, loss: 2.877716724469792e-05
step: 80, loss: 3.751329131773673e-05
step: 90, loss: 2.090216730721295e-05
step: 100, loss: 1.6379855878767557e-05
step: 110, loss: 0.0006971913389861584
step: 120, loss: 1.959091423486825e-05
step: 130, loss: 5.5596145102754235e-05
step: 140, loss: 4.4448381231632084e-05
step: 150, loss: 0.031127793714404106
step: 160, loss: 6.692843453492969e-05
step: 170, loss: 2.0365807358757593e-05
step: 180, loss: 1.7929285604623146e-05
step: 190, loss: 2.2090305719757453e-05
step: 200, loss: 2.4399805624852888e-05
step: 210, loss: 1.7635151380090974e-05
step: 220, loss: 3.1543346267426386e-05
step: 230, loss: 2.1810667021782137e-05
step: 240, loss: 0.020251121371984482
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 250, loss: 2.093198781949468e-05
step: 260, loss: 0.024874160066246986
step: 270, loss: 0.0032895884942263365
step: 280, loss: 3.781031409744173e-05
step: 290, loss: 3.1904695788398385e-05
step: 300, loss: 0.008298973552882671

```

```

step: 310, loss: 1.9728751794900745e-05
step: 320, loss: 2.6396022803965025e-05
step: 330, loss: 1.6517653421033174e-05
step: 340, loss: 1.7575623132870533e-05
step: 350, loss: 4.724190876004286e-05
step: 360, loss: 2.802427115966566e-05
step: 370, loss: 3.9887330785859376e-05
step: 380, loss: 0.0009972965344786644
step: 390, loss: 2.5900357286445796e-05
step: 400, loss: 2.7115347620565444e-05
step: 410, loss: 1.464393972128164e-05
step: 420, loss: 1.8532937247073278e-05
step: 430, loss: 0.002485696692019701
step: 440, loss: 2.7152656912221573e-05
step: 450, loss: 3.012196611962281e-05
step: 460, loss: 1.992612124013249e-05
step: 470, loss: 0.00011571931827347726
step: 480, loss: 0.00025795280816964805
step: 490, loss: 0.00010473051224835217
step: 500, loss: 0.0001120893211918883
step: 510, loss: 1.7206813936354592e-05
step: 520, loss: 1.9315220924909227e-05
step: 530, loss: 2.001188295253087e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 15: dev_f1=0.9521152952115295, f1=0.9537465309898242,
best_f1=0.9537465309898242
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02

```

```

cast_model_type      : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32   : True
master_weights       : True
loss_scale            : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
5742it [00:15, 366.48it/s]
load_f1 = 0.9510945505356311
real_f1 = 0.9483960948396094
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:11, 395.83it/s]

```

3.1.4 Amazon-Google - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
--task Structured/Amazon-Google \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \
--dk product \
--summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Amazon-Google.pt

```

```
# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/Amazon-Google \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

Selected optimization level O2: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled          : True
opt_level        : O2
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True
opt_level        : O2
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:

UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`. In PyTorch 1.1.0 and later, you should call them in the opposite order:

`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)

step: 0, loss: 0.8374627232551575

```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.6813905835151672
step: 20, loss: 0.2554158866405487
step: 30, loss: 0.2333422601222992
step: 40, loss: 0.38413509726524353
step: 50, loss: 0.43479472398757935
step: 60, loss: 0.44641542434692383
step: 70, loss: 0.19273507595062256
step: 80, loss: 0.20244231820106506
step: 90, loss: 0.3103078007698059
step: 100, loss: 0.3547845184803009
step: 110, loss: 0.2735644280910492
step: 120, loss: 0.33755558729171753
step: 130, loss: 0.1952853500843048
step: 140, loss: 0.21374735236167908
step: 150, loss: 0.08840710669755936
step: 160, loss: 0.29189711809158325
step: 170, loss: 0.2402694970369339
step: 180, loss: 0.22897684574127197
step: 190, loss: 0.23287981748580933
step: 200, loss: 0.11220742762088776
step: 210, loss: 0.15130867063999176
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.6593001841620626, f1=0.6982922201138521,
best_f1=0.6982922201138521
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09807436168193817
step: 10, loss: 0.06575619429349899
step: 20, loss: 0.1703559160232544
step: 30, loss: 0.16232351958751678
step: 40, loss: 0.25093957781791687
step: 50, loss: 0.11784466356039047
step: 60, loss: 0.19943609833717346
step: 70, loss: 0.17174144089221954
step: 80, loss: 0.14141026139259338
step: 90, loss: 0.281514436006546
step: 100, loss: 0.11077024042606354
step: 110, loss: 0.1527143120765686
step: 120, loss: 0.1398908495903015
step: 130, loss: 0.08594074845314026

```



```

step: 140, loss: 0.19453036785125732
step: 150, loss: 0.1953313946723938
step: 160, loss: 0.05545855686068535
step: 170, loss: 0.27419722080230713
step: 180, loss: 0.2179049849510193
step: 190, loss: 0.13870710134506226
step: 200, loss: 0.13636010885238647
step: 210, loss: 0.10392536967992783
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.7204968944099379, f1=0.6834381551362684,
best_f1=0.6834381551362684
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09084198623895645
step: 10, loss: 0.1771385371685028
step: 20, loss: 0.11966603994369507
step: 30, loss: 0.0688759908080101
step: 40, loss: 0.03604026883840561
step: 50, loss: 0.5630598664283752
step: 60, loss: 0.05718855932354927
step: 70, loss: 0.12110409140586853
step: 80, loss: 0.06974831223487854
step: 90, loss: 0.03779870644211769
step: 100, loss: 0.09177256375551224
step: 110, loss: 0.07085874676704407
step: 120, loss: 0.056056324392557144
step: 130, loss: 0.08267366141080856
step: 140, loss: 0.07161073386669159
step: 150, loss: 0.08667214214801788
step: 160, loss: 0.13072235882282257
step: 170, loss: 0.022553468123078346
step: 180, loss: 0.2031412422657013
step: 190, loss: 0.05986708030104637
step: 200, loss: 0.12138587236404419
step: 210, loss: 0.08678091317415237
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.

```

```

    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.7286821705426356, f1=0.6947791164658634,
best_f1=0.6947791164658634
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1511877328157425
step: 10, loss: 0.00973708275705576
step: 20, loss: 0.15544335544109344
step: 30, loss: 0.02187986485660076
step: 40, loss: 0.24696668982505798
step: 50, loss: 0.02445557713508606
step: 60, loss: 0.1885886937379837
step: 70, loss: 0.09418006241321564
step: 80, loss: 0.0729362815618515
step: 90, loss: 0.06430552899837494
step: 100, loss: 0.11439855396747589
step: 110, loss: 0.048696137964725494
step: 120, loss: 0.047748588025569916
step: 130, loss: 0.07833783328533173
step: 140, loss: 0.07487363368272781
step: 150, loss: 0.0843571126461029
step: 160, loss: 0.09423566609621048
step: 170, loss: 0.05420971289277077
step: 180, loss: 0.1112263947725296
step: 190, loss: 0.15727022290229797
step: 200, loss: 0.10026979446411133
step: 210, loss: 0.07279587537050247
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.7357293868921776, f1=0.7044025157232704,
best_f1=0.7044025157232704
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008312124758958817
step: 10, loss: 0.06815648823976517
step: 20, loss: 0.1277187317609787
step: 30, loss: 0.01026751846075058
step: 40, loss: 0.053119953721761703
step: 50, loss: 0.11061518639326096
step: 60, loss: 0.06330418586730957
step: 70, loss: 0.06326018273830414

```

```

step: 80, loss: 0.03040953166782856
step: 90, loss: 0.04758965224027634
step: 100, loss: 0.016423001885414124
step: 110, loss: 0.12648983299732208
step: 120, loss: 0.10043217986822128
step: 130, loss: 0.036058902740478516
step: 140, loss: 0.02442855015397072
step: 150, loss: 0.010081471875309944
step: 160, loss: 0.029144123196601868
step: 170, loss: 0.0337052159011364
step: 180, loss: 0.061433758586645126
step: 190, loss: 0.04961108788847923
step: 200, loss: 0.11713915318250656
step: 210, loss: 0.005869669374078512
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.7134020618556701, f1=0.6985446985446986,
best_f1=0.7044025157232704
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0693964809179306
step: 10, loss: 0.016658511012792587
step: 20, loss: 0.017359742894768715
step: 30, loss: 0.011282896623015404
step: 40, loss: 0.0015110543463379145
step: 50, loss: 0.10585454106330872
step: 60, loss: 0.013828437775373459
step: 70, loss: 0.03689063712954521
step: 80, loss: 0.031341128051280975
step: 90, loss: 0.07861055433750153
step: 100, loss: 0.008916942402720451
step: 110, loss: 0.0166252963244915
step: 120, loss: 0.015209805220365524
step: 130, loss: 0.07275311648845673
step: 140, loss: 0.036162808537483215
step: 150, loss: 0.06508852541446686
step: 160, loss: 0.07981198281049728
step: 170, loss: 0.200177863240242
step: 180, loss: 0.07588204741477966
step: 190, loss: 0.023293228819966316
step: 200, loss: 0.0344778373837471
step: 210, loss: 0.11928178369998932
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.7296296296296296, f1=0.7279549718574109,

```

```

best_f1=0.7044025157232704
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05440336838364601
step: 10, loss: 0.07453183084726334
step: 20, loss: 0.018719268962740898
step: 30, loss: 0.007283093873411417
step: 40, loss: 0.021674538031220436
step: 50, loss: 0.036394912749528885
step: 60, loss: 0.002820840571075678
step: 70, loss: 0.010655141435563564
step: 80, loss: 0.03561113029718399
step: 90, loss: 0.07891330868005753
step: 100, loss: 0.0048256972804665565
step: 110, loss: 0.013830259442329407
step: 120, loss: 0.012443766929209232
step: 130, loss: 0.02553221955895424
step: 140, loss: 0.006093541160225868
step: 150, loss: 0.014845164492726326
step: 160, loss: 0.0015533515252172947
step: 170, loss: 0.008625962771475315
step: 180, loss: 0.013844885863363743
step: 190, loss: 0.032435670495033264
step: 200, loss: 0.02203788049519062
step: 210, loss: 0.06551772356033325
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.7103174603174603, f1=0.718052738336714,
best_f1=0.7044025157232704
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000604683649726212
step: 10, loss: 0.0007247431785799563
step: 20, loss: 0.028323562815785408
step: 30, loss: 0.002794816391542554
step: 40, loss: 0.018453020602464676
step: 50, loss: 0.013923345133662224
step: 60, loss: 0.007920662872493267
step: 70, loss: 0.06277883052825928
step: 80, loss: 0.09738006442785263
step: 90, loss: 0.006554862949997187
step: 100, loss: 0.04049818217754364
step: 110, loss: 0.006740701850503683
step: 120, loss: 0.012515346519649029
step: 130, loss: 0.014653228223323822

```

```

step: 140, loss: 0.045107170939445496
step: 150, loss: 0.10920211672782898
step: 160, loss: 0.025665532797574997
step: 170, loss: 0.04339568689465523
step: 180, loss: 0.017882181331515312
step: 190, loss: 0.0005170528893359005
step: 200, loss: 0.026740988716483116
step: 210, loss: 0.054438382387161255
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.7261410788381742, f1=0.713375796178344,
best_f1=0.7044025157232704
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.045793455094099045
step: 10, loss: 0.003909267485141754
step: 20, loss: 0.004913062322884798
step: 30, loss: 0.001779167097993195
step: 40, loss: 0.0011508786119520664
step: 50, loss: 0.009938913397490978
step: 60, loss: 0.028817079961299896
step: 70, loss: 0.07068120688199997
step: 80, loss: 0.10695412755012512
step: 90, loss: 0.0005636716377921402
step: 100, loss: 0.0013053013244643807
step: 110, loss: 0.00027012452483177185
step: 120, loss: 0.0034746844321489334
step: 130, loss: 0.03077814355492592
step: 140, loss: 0.0031943735666573048
step: 150, loss: 0.0020389321725815535
step: 160, loss: 0.0005210254457779229
step: 170, loss: 0.0005800807266496122
step: 180, loss: 0.0007066678954288363
step: 190, loss: 0.0006856283289380372
step: 200, loss: 0.012896740809082985
step: 210, loss: 0.04302789270877838
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.7091633466135459, f1=0.728395061728395,
best_f1=0.7044025157232704
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2164427489042282
step: 10, loss: 0.0020248598884791136

```

```

step: 20, loss: 0.005021764896810055
step: 30, loss: 0.050910428166389465
step: 40, loss: 0.0028699429240077734
step: 50, loss: 0.02333747036755085
step: 60, loss: 0.00832434743642807
step: 70, loss: 0.021676935255527496
step: 80, loss: 0.013233521953225136
step: 90, loss: 0.049981482326984406
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 100, loss: 0.0008110883645713329
step: 110, loss: 0.09590843319892883
step: 120, loss: 0.0011723682982847095
step: 130, loss: 0.0017400060314685106
step: 140, loss: 0.0012712114257737994
step: 150, loss: 0.030843734741210938
step: 160, loss: 0.0012930496595799923
step: 170, loss: 0.0009719773661345243
step: 180, loss: 0.02291756682097912
step: 190, loss: 0.0026396624743938446
step: 200, loss: 0.004807295743376017
step: 210, loss: 0.002273936290293932
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.725, f1=0.7056367432150314, best_f1=0.7044025157232704
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0010347651550546288
step: 10, loss: 0.0006796402740292251
step: 20, loss: 0.04244005307555199
step: 30, loss: 0.0010205409489572048
step: 40, loss: 0.0010612256592139602
step: 50, loss: 0.02146160788834095
step: 60, loss: 0.0010893340222537518
step: 70, loss: 0.00030713449814356863
step: 80, loss: 0.050486572086811066
step: 90, loss: 0.01658014766871929
step: 100, loss: 0.0007132919272407889
step: 110, loss: 0.04281533136963844
step: 120, loss: 0.04607519507408142
step: 130, loss: 0.0042314715683460236
step: 140, loss: 0.09910430759191513
step: 150, loss: 0.03403306007385254
step: 160, loss: 0.03058227151632309
step: 170, loss: 0.0222484078258276
step: 180, loss: 0.037713345140218735
step: 190, loss: 0.03584584221243858

```

```

step: 200, loss: 0.011423692107200623
step: 210, loss: 0.0002690753899514675
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.7089108910891089, f1=0.7191919191919192,
best_f1=0.7044025157232704
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0014626586344093084
step: 10, loss: 0.002174748806282878
step: 20, loss: 0.0013266110327094793
step: 30, loss: 0.0031630496960133314
step: 40, loss: 0.015197977423667908
step: 50, loss: 0.02948988787829876
step: 60, loss: 0.0004595967475324869
step: 70, loss: 0.0011853716569021344
step: 80, loss: 0.023039931431412697
step: 90, loss: 0.0029111334588378668
step: 100, loss: 0.00046359715634025633
step: 110, loss: 0.018885456025600433
step: 120, loss: 0.008876202628016472
step: 130, loss: 0.08630451560020447
step: 140, loss: 0.020110346376895905
step: 150, loss: 0.02108326368033886
step: 160, loss: 0.0039422414265573025
step: 170, loss: 0.02511674351990223
step: 180, loss: 0.0061241937801241875
step: 190, loss: 0.010740337893366814
step: 200, loss: 0.016238076612353325
step: 210, loss: 0.02379067987203598
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.7213740458015266, f1=0.7421874999999999,
best_f1=0.7044025157232704
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006492062471807003
step: 10, loss: 0.0032121201511472464
step: 20, loss: 0.024522671476006508
step: 30, loss: 0.020886359736323357
step: 40, loss: 0.0014893251936882734
step: 50, loss: 0.0013312181690707803
step: 60, loss: 0.0006058010039851069
step: 70, loss: 0.007983664982020855

```

```

step: 80, loss: 0.013584529981017113
step: 90, loss: 0.0005516012315638363
step: 100, loss: 0.017738550901412964
step: 110, loss: 0.013835352845489979
step: 120, loss: 0.026629362255334854
step: 130, loss: 0.0019230709876865149
step: 140, loss: 0.10430894792079926
step: 150, loss: 0.0007471354329027236
step: 160, loss: 0.1368216574192047
step: 170, loss: 0.036369021981954575
step: 180, loss: 0.00037185216206125915
step: 190, loss: 0.0008479988900944591
step: 200, loss: 0.004985847044736147
step: 210, loss: 0.002051253104582429
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.7085020242914981, f1=0.735966735966736,
best_f1=0.7044025157232704
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0018092315876856446
step: 10, loss: 0.0014400477521121502
step: 20, loss: 0.0009958529844880104
step: 30, loss: 0.038421180099248886
step: 40, loss: 0.0006487471400760114
step: 50, loss: 0.004271307028830051
step: 60, loss: 0.0022378696594387293
step: 70, loss: 0.0008820823859423399
step: 80, loss: 0.14116090536117554
step: 90, loss: 0.022737789899110794
step: 100, loss: 0.004147634841501713
step: 110, loss: 0.007296425756067038
step: 120, loss: 0.007689966820180416
step: 130, loss: 0.009399287402629852
step: 140, loss: 0.0008836328634060919
step: 150, loss: 0.001538568176329136
step: 160, loss: 0.0009793658973649144
step: 170, loss: 0.0008593598613515496
step: 180, loss: 0.005341100972145796
step: 190, loss: 0.0008807718986645341
step: 200, loss: 0.0005062215495854616
step: 210, loss: 0.03918909654021263
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.7056451612903225, f1=0.7280163599182004,

```



```

best_f1=0.7044025157232704
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.017242731526494026
step: 10, loss: 0.0062590609304606915
step: 20, loss: 0.00022406972129829228
step: 30, loss: 0.0027467922773212194
step: 40, loss: 0.00028045219369232655
step: 50, loss: 0.003998117055743933
step: 60, loss: 0.0028492379933595657
step: 70, loss: 0.01979406364262104
step: 80, loss: 0.000376117997802794
step: 90, loss: 0.028493138030171394
step: 100, loss: 0.0007442189962603152
step: 110, loss: 0.0002684149658307433
step: 120, loss: 0.014662249013781548
step: 130, loss: 0.0004219868569634855
step: 140, loss: 0.010403773747384548
step: 150, loss: 0.0005519004771485925
step: 160, loss: 0.0001954256440512836
step: 170, loss: 0.00037052418338134885
step: 180, loss: 0.0005284872022457421
step: 190, loss: 0.00033511000219732523
step: 200, loss: 0.003369879210367799
step: 210, loss: 0.02447335608303547
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.712, f1=0.7280163599182004, best_f1=0.7044025157232704
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False

```

```

keep_batchnorm_fp32      : True
master_weights           : True
loss_scale                : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2293it [00:03, 662.05it/s]
load_f1 = 0.7164750957854407
real_f1 = 0.7083333333333334
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:10, 407.59it/s]

```

3.1.5 Walmart-Amazon - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
--task Structured/Walmart-Amazon \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \
--dk product \
--summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Walmart-Amazon.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \

```

```

--task Structured/Walmart-Amazon \
--input_path input/input_small.jsonl \
--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

warnings.warn("An input tensor was not cuda.")

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:

UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order: `optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)

step: 0, loss: 0.8559309840202332

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.1629268079996109

```

step: 20, loss: 0.15390615165233612
step: 30, loss: 0.5224305987358093
step: 40, loss: 0.26631662249565125
step: 50, loss: 0.30731290578842163
step: 60, loss: 0.3629915714263916
step: 70, loss: 0.17113231122493744
step: 80, loss: 0.5258345603942871
step: 90, loss: 0.24211502075195312
step: 100, loss: 0.21639305353164673
step: 110, loss: 0.23089855909347534
step: 120, loss: 0.41219115257263184
step: 130, loss: 0.3430738151073456
step: 140, loss: 0.30497440695762634
step: 150, loss: 0.24787627160549164
step: 160, loss: 0.1932263821363449
step: 170, loss: 0.3250133991241455
step: 180, loss: 0.25299689173698425
step: 190, loss: 0.10482152551412582
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.6181818181818182, f1=0.6538461538461539,
best_f1=0.6538461538461539
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1901882290840149
step: 10, loss: 0.028012629598379135
step: 20, loss: 0.0923677459359169
step: 30, loss: 0.1654987782239914
step: 40, loss: 0.4765363335609436
step: 50, loss: 0.3068934381008148
step: 60, loss: 0.112108513712883
step: 70, loss: 0.18774481117725372
step: 80, loss: 0.15545760095119476
step: 90, loss: 0.12339230626821518
step: 100, loss: 0.27990299463272095
step: 110, loss: 0.1301102489233017
step: 120, loss: 0.16403084993362427
step: 130, loss: 0.17754581570625305
step: 140, loss: 0.08575966954231262
step: 150, loss: 0.01048002578318119
step: 160, loss: 0.10971026122570038
step: 170, loss: 0.292196124792099

```

```

step: 180, loss: 0.2723916172981262
step: 190, loss: 0.13630113005638123
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.7513812154696132, f1=0.7603305785123966,
best_f1=0.7603305785123966
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.15911540389060974
step: 10, loss: 0.2507885992527008
step: 20, loss: 0.0971178412437439
step: 30, loss: 0.08263684064149857
step: 40, loss: 0.03238101303577423
step: 50, loss: 0.18370720744132996
step: 60, loss: 0.04087761417031288
step: 70, loss: 0.10322318226099014
step: 80, loss: 0.1258091926574707
step: 90, loss: 0.09152066707611084
step: 100, loss: 0.06572835147380829
step: 110, loss: 0.14847886562347412
step: 120, loss: 0.015497710555791855
step: 130, loss: 0.01302601583302021
step: 140, loss: 0.03672387823462486
step: 150, loss: 0.05491151660680771
step: 160, loss: 0.12612475454807281
step: 170, loss: 0.03651180490851402
step: 180, loss: 0.007449838798493147
step: 190, loss: 0.22385729849338531
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.7667560321715818, f1=0.776536312849162,
best_f1=0.776536312849162
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03367301821708679
step: 10, loss: 0.018574818968772888

```

```

step: 20, loss: 0.0077666123397648335
step: 30, loss: 0.07242776453495026
step: 40, loss: 0.0036123190075159073
step: 50, loss: 0.062237728387117386
step: 60, loss: 0.07237260788679123
step: 70, loss: 0.04453433305025101
step: 80, loss: 0.01685640960931778
step: 90, loss: 0.0820082351565361
step: 100, loss: 0.0105455806478858
step: 110, loss: 0.01460018940269947
step: 120, loss: 0.023375889286398888
step: 130, loss: 0.29825782775878906
step: 140, loss: 0.021649932488799095
step: 150, loss: 0.02038722112774849
step: 160, loss: 0.010977281257510185
step: 170, loss: 0.01339376624673605
step: 180, loss: 0.07886748015880585
step: 190, loss: 0.06686229258775711
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.7891891891891892, f1=0.7658402203856748,
best_f1=0.7658402203856748
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.053474053740501404
step: 10, loss: 0.004414868541061878
step: 20, loss: 0.026267506182193756
step: 30, loss: 0.009174604900181293
step: 40, loss: 0.004772903863340616
step: 50, loss: 0.03191891685128212
step: 60, loss: 0.01570207253098488
step: 70, loss: 0.0007358329021371901
step: 80, loss: 0.09702309221029282
step: 90, loss: 0.0008207726641558111
step: 100, loss: 0.2032088339328766
step: 110, loss: 0.013849467039108276
step: 120, loss: 0.007191204000264406
step: 130, loss: 0.005591847002506256
step: 140, loss: 0.02284732274711132
step: 150, loss: 0.019063210114836693
step: 160, loss: 0.003136351238936186
step: 170, loss: 0.007997856475412846

```

```

step: 180, loss: 0.0468158982694149
step: 190, loss: 0.14972025156021118
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.7913279132791328, f1=0.7954545454545454,
best_f1=0.7954545454545454
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11211856454610825
step: 10, loss: 0.004493802785873413
step: 20, loss: 0.0012600901536643505
step: 30, loss: 0.0013664409052580595
step: 40, loss: 0.18773017823696136
step: 50, loss: 0.0036578902509063482
step: 60, loss: 0.0007652282947674394
step: 70, loss: 0.01788177154958248
step: 80, loss: 0.006723047234117985
step: 90, loss: 0.01964893378317356
step: 100, loss: 0.009147089906036854
step: 110, loss: 0.15561619400978088
step: 120, loss: 0.004057206679135561
step: 130, loss: 0.0057606203481554985
step: 140, loss: 0.0038942203391343355
step: 150, loss: 0.023504167795181274
step: 160, loss: 0.08538711071014404
step: 170, loss: 0.1443232297897339
step: 180, loss: 0.040079228579998016
step: 190, loss: 0.05447373911738396
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.7696476964769648, f1=0.7397260273972603,
best_f1=0.7954545454545454
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0026011622976511717
step: 10, loss: 0.03605625033378601
step: 20, loss: 0.013380011543631554
step: 30, loss: 0.010226069949567318
step: 40, loss: 0.011571069248020649
step: 50, loss: 0.2350127100944519

```

```

step: 60, loss: 0.0031887362711131573
step: 70, loss: 0.03787462040781975
step: 80, loss: 0.1257934421300888
step: 90, loss: 0.011396764777600765
step: 100, loss: 0.0008679223246872425
step: 110, loss: 0.03667987138032913
step: 120, loss: 0.0020826656837016344
step: 130, loss: 0.0016825507627800107
step: 140, loss: 0.017407894134521484
step: 150, loss: 0.003361866809427738
step: 160, loss: 0.00063051882898435
step: 170, loss: 0.0007946165860630572
step: 180, loss: 0.001493337913416326
step: 190, loss: 0.002346920780837536
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.7567567567567567, f1=0.7425474254742548,
best_f1=0.7954545454545454
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006684739142656326
step: 10, loss: 0.004028875380754471
step: 20, loss: 0.0011390793370082974
step: 30, loss: 0.0009175763698294759
step: 40, loss: 0.012186095118522644
step: 50, loss: 0.06265633553266525
step: 60, loss: 0.000892867858055979
step: 70, loss: 0.0018910622457042336
step: 80, loss: 0.03285873681306839
step: 90, loss: 0.0019458190072327852
step: 100, loss: 0.008022967725992203
step: 110, loss: 0.0018257112242281437
step: 120, loss: 0.0010199870448559523
step: 130, loss: 0.010523748584091663
step: 140, loss: 0.0018967501819133759
step: 150, loss: 0.0010305597679689527
step: 160, loss: 0.0018112949328497052
step: 170, loss: 0.001032607164233923
step: 180, loss: 0.015989188104867935
step: 190, loss: 0.0026381576899439096
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.781491002570694, f1=0.7806122448979592,
best_f1=0.7954545454545454
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```


An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.0005477208760567009
step: 10, loss: 0.0038399852346628904
step: 20, loss: 0.022871946915984154
step: 30, loss: 0.001739479135721922
step: 40, loss: 0.006555576808750629
step: 50, loss: 0.0012407292379066348
step: 60, loss: 0.006346415262669325
step: 70, loss: 0.003716578707098961
step: 80, loss: 0.0022333019878715277
step: 90, loss: 0.0034853972028940916
step: 100, loss: 0.0016765162581577897
step: 110, loss: 0.0010333700338378549
step: 120, loss: 0.011957256123423576
step: 130, loss: 0.0008270156104117632
step: 140, loss: 0.001315841916948557
step: 150, loss: 0.02648879960179329
step: 160, loss: 0.008874543942511082
step: 170, loss: 0.10037320107221603
step: 180, loss: 0.002648198278620839
step: 190, loss: 0.0014542570570483804
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

```
epoch 9: dev_f1=0.7828418230563003, f1=0.7890410958904108,
best_f1=0.7954545454545454
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.0003907823993358761
step: 10, loss: 0.000398514122935012
step: 20, loss: 0.04737567901611328
step: 30, loss: 0.12220854312181473
step: 40, loss: 0.0008096566889435053
step: 50, loss: 0.0002842495741788298
step: 60, loss: 0.0007237521349452436
step: 70, loss: 0.0007132832542993128
step: 80, loss: 0.00026750273536890745
step: 90, loss: 0.01656416431069374
step: 100, loss: 0.0006183422519825399
step: 110, loss: 0.007681278511881828
step: 120, loss: 0.005526028573513031
step: 130, loss: 0.0011514229699969292
step: 140, loss: 0.04097219929099083
step: 150, loss: 0.00166179402731359
step: 160, loss: 0.000742295291274786
step: 170, loss: 0.05230332911014557
```

```

step: 180, loss: 0.006717987824231386
step: 190, loss: 0.009170685894787312
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 10: dev_f1=0.7967914438502673, f1=0.772972972972973,
best_f1=0.772972972972973
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000296516198432073
step: 10, loss: 0.0006520027527585626
step: 20, loss: 0.0010322908638045192
step: 30, loss: 0.00369644770398736
step: 40, loss: 0.0003176751488354057
step: 50, loss: 0.0009358247625641525
step: 60, loss: 0.006683585699647665
step: 70, loss: 0.0001940684305736795
step: 80, loss: 0.00035834216396324337
step: 90, loss: 0.0020562943536788225
step: 100, loss: 0.00023228401551023126
step: 110, loss: 0.0002384421823080629
step: 120, loss: 0.001445234869606793
step: 130, loss: 0.0002615592966321856
step: 140, loss: 0.0005316773895174265
step: 150, loss: 0.00040724084828980267
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 160, loss: 0.0003686430864036083
step: 170, loss: 0.0028880671598017216
step: 180, loss: 0.008169472217559814
step: 190, loss: 0.0003486669738776982
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.7821522309711286, f1=0.7810026385224274,
best_f1=0.772972972972973
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002625433844514191
step: 10, loss: 0.0006322340341284871
step: 20, loss: 0.0005967729957774282
step: 30, loss: 0.002845768118277192
step: 40, loss: 0.0017284955829381943

```

```

step: 50, loss: 0.0093460064381361
step: 60, loss: 0.00032161048147827387
step: 70, loss: 0.0007744731265120208
step: 80, loss: 0.0004739801515825093
step: 90, loss: 0.0004423052305355668
step: 100, loss: 0.0004003307258244604
step: 110, loss: 0.0002717569295782596
step: 120, loss: 0.00021983537590131164
step: 130, loss: 0.00044343585614115
step: 140, loss: 0.0028117825277149677
step: 150, loss: 0.00033983238972723484
step: 160, loss: 0.0004815112624783069
step: 170, loss: 0.000988698797300458
step: 180, loss: 0.00028262511477805674
step: 190, loss: 0.0008880752138793468
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.7667560321715818, f1=0.7798408488063661,
best_f1=0.772972972972973
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00022433471167460084
step: 10, loss: 0.0002112693910021335
step: 20, loss: 0.0011731797130778432
step: 30, loss: 0.0005606264458037913
step: 40, loss: 0.00030488622724078596
step: 50, loss: 0.0003136004670523107
step: 60, loss: 0.00016491321730427444
step: 70, loss: 0.0010982800740748644
step: 80, loss: 0.00047517562052235007
step: 90, loss: 0.00020987070456612855
step: 100, loss: 0.0005877283401787281
step: 110, loss: 0.0007152602775022388
step: 120, loss: 0.0003432647790759802
step: 130, loss: 0.007734025828540325
step: 140, loss: 0.009854189120233059
step: 150, loss: 0.0005162499728612602
step: 160, loss: 0.0002405670820735395
step: 170, loss: 0.0239627193659544
step: 180, loss: 0.00621280912309885
step: 190, loss: 0.00043698810623027384
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.7727272727272727, f1=0.7670454545454546,
best_f1=0.772972972972973

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00040963993524201214
step: 10, loss: 0.0003693989128805697
step: 20, loss: 0.0003319616662338376
step: 30, loss: 0.0003211361763533205
step: 40, loss: 0.0008816947229206562
step: 50, loss: 0.14540672302246094
step: 60, loss: 0.00043722399277612567
step: 70, loss: 0.0005979708512313664
step: 80, loss: 0.0001774040429154411
step: 90, loss: 0.0003115767613053322
step: 100, loss: 0.00029869144782423973
step: 110, loss: 0.0003340887778904289
step: 120, loss: 0.0007441188208758831
step: 130, loss: 0.0002132769877789542
step: 140, loss: 0.00025519949849694967
step: 150, loss: 0.0007565035484731197
step: 160, loss: 0.00034777665860019624
step: 170, loss: 0.0017410429427400231
step: 180, loss: 0.00040647920104674995
step: 190, loss: 0.002008285606279969
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.774928774928775, f1=0.7758620689655173,
best_f1=0.772972972972973
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00042916531674563885
step: 10, loss: 0.00022516740136779845
step: 20, loss: 0.00023474785848520696
step: 30, loss: 0.0004948829882778227
step: 40, loss: 0.00032671241206116974
step: 50, loss: 0.0014715477591380477
step: 60, loss: 0.0003203683299943805
step: 70, loss: 0.0004418124444782734
step: 80, loss: 0.0017343283398076892
step: 90, loss: 0.004053372424095869
step: 100, loss: 0.00027658752514980733
step: 110, loss: 0.0007063389057293534
step: 120, loss: 0.0083417734131217
step: 130, loss: 0.000330588489305228
step: 140, loss: 0.002379031153395772
step: 150, loss: 0.00046318041859194636
step: 160, loss: 0.002023387234658003

```

```

step: 170, loss: 0.0004283690359443426
step: 180, loss: 0.0005725740338675678
step: 190, loss: 0.0003710939490702003
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.779220779220779, f1=0.772845953002611,
best_f1=0.772972972972973
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2049it [00:09, 225.30it/s]
load_f1 = 0.6052631578947368
real_f1 = 0.6081081081081081
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:18, 244.27it/s]

```

3.2 DK TEXTUAL

3.2.1 Abt-Buy - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
--task Textual/Abt-Buy \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \
--dk product \
--summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Abt-Buy.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
--task Textual/Abt-Buy \
--input_path input/input_small.jsonl \
--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

enabled	: True
opt_level	: 02
cast_model_type	: torch.float16
patch_torch_functions	: False
keep_batchnorm_fp32	: True
master_weights	: True
loss_scale	: dynamic

```

Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale              : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8546394109725952
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.22392511367797852
step: 20, loss: 0.15466946363449097
step: 30, loss: 0.23219335079193115
step: 40, loss: 0.3105788826942444
step: 50, loss: 0.3729967176914215
step: 60, loss: 0.43948376178741455
step: 70, loss: 0.3146539628505707
step: 80, loss: 0.2534307539463043
step: 90, loss: 0.3988999128341675
step: 100, loss: 0.23487287759780884
step: 110, loss: 0.17801688611507416
step: 120, loss: 0.5208708643913269
step: 130, loss: 0.371229887008667
step: 140, loss: 0.35574573278427124
step: 150, loss: 0.10029995441436768
step: 160, loss: 0.22309240698814392
step: 170, loss: 0.13652338087558746
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:

```

UserWarning: Please also save or load the state of the optimizer when saving or loading the scheduler.

```
warnings.warn(SAVE_STATE_WARNING, UserWarning)
```

epoch 1: dev_f1=0.7238095238095237, f1=0.7064220183486238,

best_f1=0.7064220183486238

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

step: 0, loss: 0.1431020200252533

step: 10, loss: 0.03167823702096939

step: 20, loss: 0.28602084517478943

step: 30, loss: 0.046002347022295

step: 40, loss: 0.1686154305934906

step: 50, loss: 0.23045563697814941

step: 60, loss: 0.03821643814444542

step: 70, loss: 0.17820529639720917

step: 80, loss: 0.16067321598529816

step: 90, loss: 0.16101230680942535

step: 100, loss: 0.12056783586740494

step: 110, loss: 0.16927973926067352

step: 120, loss: 0.0350630097091198

step: 130, loss: 0.06779433786869049

step: 140, loss: 0.1259188950061798

step: 150, loss: 0.06048369035124779

step: 160, loss: 0.0387851744890213

step: 170, loss: 0.04530596733093262

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:

UserWarning: Please also save or load the state of the optimizer when saving or loading the scheduler.

```
warnings.warn(SAVE_STATE_WARNING, UserWarning)
```

epoch 2: dev_f1=0.7890818858560793, f1=0.754257907542579,

best_f1=0.754257907542579

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

step: 0, loss: 0.05657607316970825

step: 10, loss: 0.013074154034256935

step: 20, loss: 0.12624631822109222

step: 30, loss: 0.08462730795145035

step: 40, loss: 0.012570326216518879

step: 50, loss: 0.17249248921871185

step: 60, loss: 0.05306226387619972

step: 70, loss: 0.017729636281728745

step: 80, loss: 0.1368527114391327

step: 90, loss: 0.094178706407547


```

step: 100, loss: 0.02289082482457161
step: 110, loss: 0.08377144485712051
step: 120, loss: 0.1145695298910141
step: 130, loss: 0.12703078985214233
step: 140, loss: 0.006145186722278595
step: 150, loss: 0.01393884140998125
step: 160, loss: 0.23742888867855072
step: 170, loss: 0.13320139050483704
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.7713625866050808, f1=0.7578475336322871,
best_f1=0.754257907542579
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07737063616514206
step: 10, loss: 0.1126130148768425
step: 20, loss: 0.0421108677983284
step: 30, loss: 0.030625952407717705
step: 40, loss: 0.008373240008950233
step: 50, loss: 0.008559654466807842
step: 60, loss: 0.11675417423248291
step: 70, loss: 0.00710453512147069
step: 80, loss: 0.10806671530008316
step: 90, loss: 0.024619698524475098
step: 100, loss: 0.020357556641101837
step: 110, loss: 0.02938958816230297
step: 120, loss: 0.1631668210029602
step: 130, loss: 0.013797867111861706
step: 140, loss: 0.011199328117072582
step: 150, loss: 0.019480539485812187
step: 160, loss: 0.021099358797073364
step: 170, loss: 0.011354722082614899
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.8038277511961723, f1=0.7834101382488479,
best_f1=0.7834101382488479
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009967261925339699
step: 10, loss: 0.039538703858852386

```

```

step: 20, loss: 0.16531077027320862
step: 30, loss: 0.03932758793234825
step: 40, loss: 0.0030690899584442377
step: 50, loss: 0.00491579482331872
step: 60, loss: 0.0054094516672194
step: 70, loss: 0.010434497147798538
step: 80, loss: 0.021237622946500778
step: 90, loss: 0.015759097412228584
step: 100, loss: 0.02554207295179367
step: 110, loss: 0.007918717339634895
step: 120, loss: 0.05544089898467064
step: 130, loss: 0.008016971871256828
step: 140, loss: 0.0422053337097168
step: 150, loss: 0.013083189725875854
step: 160, loss: 0.010705125518143177
step: 170, loss: 0.028324434533715248
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.800982800982801, f1=0.7952941176470588,
best_f1=0.7834101382488479
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.056102100759744644
step: 10, loss: 0.06768222898244858
step: 20, loss: 0.010027033276855946
step: 30, loss: 0.04840979352593422
step: 40, loss: 0.012836252339184284
step: 50, loss: 0.004350653383880854
step: 60, loss: 0.11467241495847702
step: 70, loss: 0.007624855265021324
step: 80, loss: 0.15123897790908813
step: 90, loss: 0.07853227853775024
step: 100, loss: 0.061510760337114334
step: 110, loss: 0.006574380211532116
step: 120, loss: 0.006441507022827864
step: 130, loss: 0.1009083017706871
step: 140, loss: 0.1922328919172287
step: 150, loss: 0.13208451867103577
step: 160, loss: 0.15981560945510864
step: 170, loss: 0.008617421612143517
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.7880184331797235, f1=0.7937915742793792,
best_f1=0.7834101382488479
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009346677921712399
step: 10, loss: 0.00244052498601377
step: 20, loss: 0.19503943622112274
step: 30, loss: 0.0035403473302721977
step: 40, loss: 0.19189520180225372
step: 50, loss: 0.001682076370343566
step: 60, loss: 0.2616480588912964
step: 70, loss: 0.010813726112246513
step: 80, loss: 0.0025341373402625322
step: 90, loss: 0.004308686591684818
step: 100, loss: 0.08497513085603714
step: 110, loss: 0.0008364152163267136
step: 120, loss: 0.14704476296901703
step: 130, loss: 0.04981040954589844
step: 140, loss: 0.012900754809379578
step: 150, loss: 0.03334108740091324
step: 160, loss: 0.005585950333625078
step: 170, loss: 0.2904643416404724
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.7880434782608696, f1=0.804177545691906,
best_f1=0.7834101382488479
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0108496630564332
step: 10, loss: 0.003984491806477308
step: 20, loss: 0.07510827481746674
step: 30, loss: 0.002382833044975996
step: 40, loss: 0.009303448721766472
step: 50, loss: 0.003927914425730705
step: 60, loss: 0.02545659802854061
step: 70, loss: 0.013561414554715157
step: 80, loss: 0.012276045978069305
step: 90, loss: 0.011258476413786411
step: 100, loss: 0.003558480879291892
step: 110, loss: 0.03655153885483742
step: 120, loss: 0.0022242343984544277
step: 130, loss: 0.0019060330232605338
step: 140, loss: 0.003275969298556447
step: 150, loss: 0.03966881334781647
step: 160, loss: 0.017711691558361053
step: 170, loss: 0.0047446186654269695
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.8018648018648018, f1=0.8036529680365296,
best_f1=0.7834101382488479
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0038571658078581095
step: 10, loss: 0.007672335021197796
step: 20, loss: 0.0010103550739586353
step: 30, loss: 0.009272383525967598
step: 40, loss: 0.008312092162668705
step: 50, loss: 0.025670748203992844
step: 60, loss: 0.0076429639011621475
step: 70, loss: 0.05285482108592987
step: 80, loss: 0.09345706552267075
step: 90, loss: 0.0034170045983046293
step: 100, loss: 0.022882932797074318
step: 110, loss: 0.03308366611599922
step: 120, loss: 0.002683012979105115
step: 130, loss: 0.0021936867851763964
step: 140, loss: 0.0012283981777727604
step: 150, loss: 0.022346267476677895
step: 160, loss: 0.049897488206624985
step: 170, loss: 0.04635259509086609
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.7928388746803069, f1=0.8088235294117646,
best_f1=0.7834101382488479
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0033467288594692945
step: 10, loss: 0.06166667863726616
step: 20, loss: 0.0005084258154965937
step: 30, loss: 0.0005466713919304311
step: 40, loss: 0.0008662639884278178
step: 50, loss: 0.011789130046963692
step: 60, loss: 0.0034471426624804735
step: 70, loss: 0.0020983917638659477
step: 80, loss: 0.004714632406830788
step: 90, loss: 0.11865857243537903
step: 100, loss: 0.0010580631205812097
step: 110, loss: 0.00043121539056301117
step: 120, loss: 0.014893791638314724
step: 130, loss: 0.013698424212634563
step: 140, loss: 0.018481530249118805
step: 150, loss: 0.028094187378883362

```

```

step: 160, loss: 0.00019304404850117862
step: 170, loss: 0.0016701746499165893
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 10: dev_f1=0.8088235294117646, f1=0.815165876777251,
best_f1=0.815165876777251
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000447483645984903
step: 10, loss: 0.031359557062387466
step: 20, loss: 0.01239477563649416
step: 30, loss: 0.01481898594647646
step: 40, loss: 0.00502955773845315
step: 50, loss: 0.0008598135900683701
step: 60, loss: 0.08018822968006134
step: 70, loss: 0.010145473293960094
step: 80, loss: 0.0006084751221351326
step: 90, loss: 0.0054901172406971455
step: 100, loss: 0.0002817478380165994
step: 110, loss: 0.0016654073260724545
step: 120, loss: 0.017068639397621155
step: 130, loss: 0.0012350521283224225
step: 140, loss: 0.027415873482823372
step: 150, loss: 0.003108008299022913
step: 160, loss: 0.008169549517333508
step: 170, loss: 0.05236842483282089
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.7858942065491185, f1=0.7980769230769231,
best_f1=0.815165876777251
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01872396096587181
step: 10, loss: 0.00581512413918972
step: 20, loss: 0.034023527055978775
step: 30, loss: 0.0059370738454163074
step: 40, loss: 0.0004291908990126103
step: 50, loss: 0.0009785271249711514
step: 60, loss: 0.0016048136167228222
step: 70, loss: 0.007787538692355156

```

```

step: 80, loss: 0.020634850487113
step: 90, loss: 0.0018393639475107193
step: 100, loss: 0.00022748028277419508
step: 110, loss: 0.0005473017226904631
step: 120, loss: 0.0003913275431841612
step: 130, loss: 0.00033847559825517237
step: 140, loss: 0.000526583578903228
step: 150, loss: 0.003965992946177721
step: 160, loss: 0.0019173265900462866
step: 170, loss: 0.08266335725784302
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.7868020304568528, f1=0.8087167070217919,
best_f1=0.815165876777251
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006340252002701163
step: 10, loss: 0.059733543545007706
step: 20, loss: 0.07450836896896362
step: 30, loss: 0.0003213540476281196
step: 40, loss: 0.00040589962736703455
step: 50, loss: 0.00227075582370162
step: 60, loss: 0.0029371443670243025
step: 70, loss: 0.005923362448811531
step: 80, loss: 0.002924300730228424
step: 90, loss: 0.0002584046742413193
step: 100, loss: 0.00016816282004583627
step: 110, loss: 0.0005819241632707417
step: 120, loss: 0.0011178820859640837
step: 130, loss: 0.011386181227862835
step: 140, loss: 0.0015472550876438618
step: 150, loss: 0.00018306112906429917
step: 160, loss: 0.022651132196187973
step: 170, loss: 0.006734275259077549
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.7989556135770236, f1=0.8129675810473815,
best_f1=0.815165876777251
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006211712025105953
step: 10, loss: 0.0005557553959079087
step: 20, loss: 0.017700647935271263
step: 30, loss: 0.0026649257633835077

```

```

step: 40, loss: 0.00235143699683249
step: 50, loss: 0.0024144488852471113
step: 60, loss: 0.0027123864274472
step: 70, loss: 0.009023458696901798
step: 80, loss: 0.0041487254202365875
step: 90, loss: 0.0030290381982922554
step: 100, loss: 0.0012331203324720263
step: 110, loss: 0.0003010523214470595
step: 120, loss: 0.0014877067878842354
step: 130, loss: 0.0004266277246642858
step: 140, loss: 0.001012783613987267
step: 150, loss: 0.000993172638118267
step: 160, loss: 0.017817337065935135
step: 170, loss: 0.0010204887948930264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.7967914438502675, f1=0.8140703517587939,
best_f1=0.815165876777251
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002505944110453129
step: 10, loss: 0.0038127072621136904
step: 20, loss: 0.0019131341250613332
step: 30, loss: 0.01836935430765152
step: 40, loss: 0.0002632440300658345
step: 50, loss: 0.006201432552188635
step: 60, loss: 0.00011618712596828118
step: 70, loss: 0.0008774264715611935
step: 80, loss: 0.0003266960266046226
step: 90, loss: 0.0002743372169788927
step: 100, loss: 0.11434409767389297
step: 110, loss: 0.0001677368418313563
step: 120, loss: 0.02808661200106144
step: 130, loss: 0.002131829271093011
step: 140, loss: 0.001349258003756404
step: 150, loss: 0.03254465386271477
step: 160, loss: 0.00019913309370167553
step: 170, loss: 0.0021366875153034925
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.798941798941799, f1=0.8140703517587939,
best_f1=0.815165876777251
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

```

```

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
1916it [00:05, 330.70it/s]
load_f1 = 0.6860158311345647
real_f1 = 0.6666666666666667
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:17, 253.29it/s]

```

3.3 DK DIRTY

3.3.1 DBLP-ACM - Running the matcher

```

[ ]: # Step 1:
      # train the wdc model with the save_model flag on
      # it will produce two files *_dev.pt and *_test.pt which are the
      # best checkpoints on the dev set and the test set.
      !CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
        --task Dirty/DBLP-ACM \
        --batch_size 32 \

```



```

--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \
--dk product \
--summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-ACM.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Dirty/DBLP-ACM \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.

```

Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.7974493503570557
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.45322149991989136
step: 20, loss: 0.5647850632667542
step: 30, loss: 0.4612817168235779
step: 40, loss: 0.2614852488040924
step: 50, loss: 0.1728358119726181
step: 60, loss: 0.044600144028663635
step: 70, loss: 0.18547391891479492
step: 80, loss: 0.21476741135120392
step: 90, loss: 0.05551869422197342
step: 100, loss: 0.058313872665166855
step: 110, loss: 0.12827081978321075
step: 120, loss: 0.048518117517232895
step: 130, loss: 0.037432316690683365
step: 140, loss: 0.025412360206246376
step: 150, loss: 0.17808978259563446
step: 160, loss: 0.21470431983470917
step: 170, loss: 0.038209445774555206
step: 180, loss: 0.03817267343401909
step: 190, loss: 0.011705228127539158
step: 200, loss: 0.016835326328873634
step: 210, loss: 0.02509399503469467
step: 220, loss: 0.018277902156114578
step: 230, loss: 0.01999429054558277
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9503858875413451, f1=0.9553571428571428,
best_f1=0.9553571428571428

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11852476000785828
step: 10, loss: 0.027537411078810692
step: 20, loss: 0.004556210711598396
step: 30, loss: 0.00369024695828557
step: 40, loss: 0.0030057867988944054
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 50, loss: 0.023068498820066452
step: 60, loss: 0.0038451862055808306
step: 70, loss: 0.06457661092281342
step: 80, loss: 0.007521855179220438
step: 90, loss: 0.02022670954465866
step: 100, loss: 0.13184434175491333
step: 110, loss: 0.12966962158679962
step: 120, loss: 0.0067835962399840355
step: 130, loss: 0.016019757837057114
step: 140, loss: 0.29984626173973083
step: 150, loss: 0.020763318985700607
step: 160, loss: 0.011535748839378357
step: 170, loss: 0.019781114533543587
step: 180, loss: 0.008566745556890965
step: 190, loss: 0.09194077551364899
step: 200, loss: 0.019718967378139496
step: 210, loss: 0.12563878297805786
step: 220, loss: 0.006116824224591255
step: 230, loss: 0.02796262502670288
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.967525195968645, f1=0.9665178571428571,
best_f1=0.9665178571428571
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10557489097118378
step: 10, loss: 0.06649941205978394
step: 20, loss: 0.008237808011472225
step: 30, loss: 0.11000300198793411
step: 40, loss: 0.06746652722358704
step: 50, loss: 0.005860630888491869
step: 60, loss: 0.035728808492422104
step: 70, loss: 0.008237491361796856

```

```

step: 80, loss: 0.0032231134828180075
step: 90, loss: 0.08102186769247055
step: 100, loss: 0.012205466628074646
step: 110, loss: 0.00816617626696825
step: 120, loss: 0.00621531717479229
step: 130, loss: 0.0030015858355909586
step: 140, loss: 0.0049408674240112305
step: 150, loss: 0.0078004165552556515
step: 160, loss: 0.010540267452597618
step: 170, loss: 0.0033474720548838377
step: 180, loss: 0.003597402013838291
step: 190, loss: 0.005014647264033556
step: 200, loss: 0.009382291696965694
step: 210, loss: 0.013099424540996552
step: 220, loss: 0.02301434427499771
step: 230, loss: 0.05097223073244095
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.9644444444444443, f1=0.9589345172031077,
best_f1=0.9665178571428571
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010010385885834694
step: 10, loss: 0.0005394503241404891
step: 20, loss: 0.0021472834050655365
step: 30, loss: 0.0004402974445838481
step: 40, loss: 0.0016761624719947577
step: 50, loss: 0.00225722580216825
step: 60, loss: 0.0022877263836562634
step: 70, loss: 0.03831958398222923
step: 80, loss: 0.13646425306797028
step: 90, loss: 0.006953919772058725
step: 100, loss: 0.0028813884127885103
step: 110, loss: 0.056856412440538406
step: 120, loss: 0.010794198140501976
step: 130, loss: 0.005480403546243906
step: 140, loss: 0.00029552800697274506
step: 150, loss: 0.0016965671675279737
step: 160, loss: 0.0013181599788367748
step: 170, loss: 0.013180549256503582
step: 180, loss: 0.10787422209978104
step: 190, loss: 0.009620760567486286
step: 200, loss: 0.003400546731427312
step: 210, loss: 0.0024080153089016676
step: 220, loss: 0.001227551605552435
step: 230, loss: 0.0005713980062864721

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.9696969696969697, f1=0.9707865168539327,
best_f1=0.9707865168539327
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005090166232548654
step: 10, loss: 0.0030146108474582434
step: 20, loss: 0.0004353043041191995
step: 30, loss: 0.00043782632565125823
step: 40, loss: 0.0004722205630969256
step: 50, loss: 0.0015027073677629232
step: 60, loss: 0.001169272349216044
step: 70, loss: 0.00021227795514278114
step: 80, loss: 0.0014944435097277164
step: 90, loss: 0.0002938915276899934
step: 100, loss: 0.004018099512904882
step: 110, loss: 0.0003450984077062458
step: 120, loss: 0.05734098330140114
step: 130, loss: 0.022894471883773804
step: 140, loss: 0.000987864681519568
step: 150, loss: 0.00039209285750985146
step: 160, loss: 0.1203727275133133
step: 170, loss: 0.22692376375198364
step: 180, loss: 0.002459992654621601
step: 190, loss: 0.007690373808145523
step: 200, loss: 0.002260139212012291
step: 210, loss: 0.000514552288223058
step: 220, loss: 0.002142333658412099
step: 230, loss: 0.0008725400548428297
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.967452300785634, f1=0.96875, best_f1=0.9707865168539327
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00039007276063784957
step: 10, loss: 0.0003840481804218143
step: 20, loss: 0.0006359635735861957
step: 30, loss: 0.00024744425900280476
step: 40, loss: 0.0006789191975258291

```

```

step: 50, loss: 0.025180373340845108
step: 60, loss: 0.0436953529715538
step: 70, loss: 0.0007315655238926411
step: 80, loss: 0.0005201632739044726
step: 90, loss: 0.002752111293375492
step: 100, loss: 0.0012152674607932568
step: 110, loss: 0.03345257416367531
step: 120, loss: 0.0008836127235554159
step: 130, loss: 0.005368759389966726
step: 140, loss: 0.0010594280902296305
step: 150, loss: 0.0009060007869265974
step: 160, loss: 0.005827086511999369
step: 170, loss: 0.0021389645989984274
step: 180, loss: 0.002894807606935501
step: 190, loss: 0.0002796512853819877
step: 200, loss: 0.0004774243861902505
step: 210, loss: 0.0012435049284249544
step: 220, loss: 0.0014430865412577987
step: 230, loss: 0.005801615305244923
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.967305524239008, f1=0.9730337078651685,
best_f1=0.9707865168539327
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05754444748163223
step: 10, loss: 0.002241014502942562
step: 20, loss: 0.001790364971384406
step: 30, loss: 0.0006864074966870248
step: 40, loss: 0.0034357646945863962
step: 50, loss: 0.00015591035480611026
step: 60, loss: 0.0121120885014534
step: 70, loss: 0.00031116610625758767
step: 80, loss: 0.00033169021480716765
step: 90, loss: 0.002527287695556879
step: 100, loss: 0.003386894939467311
step: 110, loss: 0.006035695318132639
step: 120, loss: 0.000418908050050959
step: 130, loss: 0.0013643164420500398
step: 140, loss: 0.00013861339539289474
step: 150, loss: 0.08186082541942596
step: 160, loss: 0.0001280013530049473
step: 170, loss: 0.0003985503863077611
step: 180, loss: 0.000303921930026263
step: 190, loss: 0.0002951506176032126
step: 200, loss: 0.0003367207245901227

```

```

step: 210, loss: 0.0008931939955800772
step: 220, loss: 0.01234169490635395
step: 230, loss: 0.045951854437589645
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.9698324022346367, f1=0.9665924276169264,
best_f1=0.9665924276169264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.028473209589719772
step: 10, loss: 0.016462767496705055
step: 20, loss: 0.0003364885051269084
step: 30, loss: 0.0002308404364157468
step: 40, loss: 0.0015917217824608088
step: 50, loss: 0.02848859131336212
step: 60, loss: 0.000760024762712419
step: 70, loss: 0.0006717692012898624
step: 80, loss: 0.003168330993503332
step: 90, loss: 0.00014397819177247584
step: 100, loss: 0.0004414325230754912
step: 110, loss: 0.0012104945490136743
step: 120, loss: 0.0012514784466475248
step: 130, loss: 0.00022931158309802413
step: 140, loss: 8.882548718247563e-05
step: 150, loss: 0.00012165067164460197
step: 160, loss: 0.0005479610408656299
step: 170, loss: 0.0004073993768543005
step: 180, loss: 0.00028760937857441604
step: 190, loss: 5.366711411625147e-05
step: 200, loss: 0.00020628170750569552
step: 210, loss: 0.00012966443318873644
step: 220, loss: 0.00032437380286864936
step: 230, loss: 0.008091738447546959
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9684684684684683, f1=0.9718785151856018,
best_f1=0.9665924276169264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0011937407543882728

```

```

step: 10, loss: 0.0018850284395739436
step: 20, loss: 0.007568332366645336
step: 30, loss: 0.0037368424236774445
step: 40, loss: 0.0001353526022285223
step: 50, loss: 0.00010081196523969993
step: 60, loss: 0.00030815263744443655
step: 70, loss: 0.000108220090623945
step: 80, loss: 0.05293875187635422
step: 90, loss: 0.09440163522958755
step: 100, loss: 0.006812380161136389
step: 110, loss: 0.0001319329021498561
step: 120, loss: 0.06656649708747864
step: 130, loss: 0.00018297410861123353
step: 140, loss: 0.00569231016561389
step: 150, loss: 0.00010524561366764829
step: 160, loss: 0.0013351234374567866
step: 170, loss: 0.0001201145350933075
step: 180, loss: 0.0078693563118577
step: 190, loss: 0.00016149228031281382
step: 200, loss: 0.00029787656967528164
step: 210, loss: 0.00014322805509436876
step: 220, loss: 0.09106677770614624
step: 230, loss: 0.027053935453295708
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9665178571428571, f1=0.967741935483871,
best_f1=0.9665924276169264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00015469896607100964
step: 10, loss: 4.9416456022299826e-05
step: 20, loss: 0.00015851645730435848
step: 30, loss: 0.00023638451239094138
step: 40, loss: 0.001839260570704937
step: 50, loss: 0.00014762183127459139
step: 60, loss: 0.07459861785173416
step: 70, loss: 8.49632706376724e-05
step: 80, loss: 0.00011451861792011186
step: 90, loss: 8.138988050632179e-05
step: 100, loss: 8.25661700218916e-05
step: 110, loss: 9.511740790912881e-05
step: 120, loss: 0.015383881516754627
step: 130, loss: 0.016680069267749786
step: 140, loss: 0.05506616458296776
step: 150, loss: 9.544753993395716e-05
step: 160, loss: 0.000279921485343948

```



```

step: 170, loss: 0.0015563678462058306
step: 180, loss: 0.0001289766514673829
step: 190, loss: 0.0012118399608880281
step: 200, loss: 4.4407574023352936e-05
step: 210, loss: 4.5859909732826054e-05
step: 220, loss: 0.00021014644880779088
step: 230, loss: 0.00014401199587155133
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9659090909090909, f1=0.9694915254237287,
best_f1=0.9665924276169264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.2087707445025444e-05
step: 10, loss: 0.00017036657663993537
step: 20, loss: 5.337956827133894e-05
step: 30, loss: 0.00017361763457302004
step: 40, loss: 0.002806547097861767
step: 50, loss: 8.919322863221169e-05
step: 60, loss: 0.0010144609259441495
step: 70, loss: 7.004545477684587e-05
step: 80, loss: 8.285920193884522e-05
step: 90, loss: 0.00024543903418816626
step: 100, loss: 8.384750981349498e-05
step: 110, loss: 7.230389019241557e-05
step: 120, loss: 0.0011863502440974116
step: 130, loss: 5.899997995584272e-05
step: 140, loss: 4.055897079524584e-05
step: 150, loss: 4.439126132638194e-05
step: 160, loss: 0.00010035824379883707
step: 170, loss: 0.0017232417594641447
step: 180, loss: 0.00011497087689349428
step: 190, loss: 0.002550970297306776
step: 200, loss: 5.751560820499435e-05
step: 210, loss: 5.2754334319615737e-05
step: 220, loss: 4.625856308848597e-05
step: 230, loss: 7.270940841408446e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9659090909090909, f1=0.971687429218573,
best_f1=0.9665924276169264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00016806540952529758

```

```

step: 10, loss: 0.0008521285490132868
step: 20, loss: 4.8479745601071045e-05
step: 30, loss: 0.0004088407731615007
step: 40, loss: 3.654396641650237e-05
step: 50, loss: 9.43285267567262e-05
step: 60, loss: 0.006561690010130405
step: 70, loss: 5.4789928981335834e-05
step: 80, loss: 4.7944799007382244e-05
step: 90, loss: 5.435439743450843e-05
step: 100, loss: 8.349159179488197e-05
step: 110, loss: 6.996899173827842e-05
step: 120, loss: 5.488498572958633e-05
step: 130, loss: 6.513559492304921e-05
step: 140, loss: 4.7628596803406253e-05
step: 150, loss: 0.00013533640594687313
step: 160, loss: 8.355681347893551e-05
step: 170, loss: 6.019955253577791e-05
step: 180, loss: 0.00038097691140137613
step: 190, loss: 5.6450822739861906e-05
step: 200, loss: 0.0018927013734355569
step: 210, loss: 5.19119857926853e-05
step: 220, loss: 0.00013470303383655846
step: 230, loss: 9.484411566518247e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9679558011049725, f1=0.9668141592920354,
best_f1=0.9665924276169264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000533281359821558
step: 10, loss: 0.00018475882825441658
step: 20, loss: 9.103392949327826e-05
step: 30, loss: 0.008393827825784683
step: 40, loss: 0.0026041632518172264
step: 50, loss: 0.00013428983220364898
step: 60, loss: 6.196348113007843e-05
step: 70, loss: 6.976736767683178e-05
step: 80, loss: 4.5690136175835505e-05
step: 90, loss: 5.821741433464922e-05
step: 100, loss: 6.220796058187261e-05
step: 110, loss: 6.232709711184725e-05
step: 120, loss: 0.00014811565051786602
step: 130, loss: 8.144557068590075e-05
step: 140, loss: 4.558336877380498e-05
step: 150, loss: 0.00012633921869564801
step: 160, loss: 6.063358159735799e-05

```

```

step: 170, loss: 7.341608579736203e-05
step: 180, loss: 0.0001760332379490137
step: 190, loss: 3.5020115319639444e-05
step: 200, loss: 4.441029886947945e-05
step: 210, loss: 8.689833339303732e-05
step: 220, loss: 0.00011006633576471359
step: 230, loss: 5.0122565880883485e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9666666666666666, f1=0.9633740288568259,
best_f1=0.9665924276169264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.657722820411436e-05
step: 10, loss: 3.3864733268273994e-05
step: 20, loss: 0.015615785494446754
step: 30, loss: 0.00048129193601198494
step: 40, loss: 3.544070204952732e-05
step: 50, loss: 0.0027779648080468178
step: 60, loss: 5.966014941805042e-05
step: 70, loss: 5.552858783630654e-05
step: 80, loss: 6.517111614812165e-05
step: 90, loss: 8.653346594655886e-05
step: 100, loss: 0.015711713582277298
step: 110, loss: 0.0002495532389730215
step: 120, loss: 0.0030418802052736282
step: 130, loss: 5.1484748837538064e-05
step: 140, loss: 0.00011896203795913607
step: 150, loss: 5.1913018978666514e-05
step: 160, loss: 2.9171729693189263e-05
step: 170, loss: 5.9310033975634724e-05
step: 180, loss: 7.557539356639609e-05
step: 190, loss: 0.0008099651895463467
step: 200, loss: 0.00029217396513558924
step: 210, loss: 0.012101901695132256
step: 220, loss: 0.0007496651960536838
step: 230, loss: 2.5316476239822805e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9678135405105438, f1=0.9655172413793103,
best_f1=0.9665924276169264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.3533855457790196e-05

```

```

step: 10, loss: 0.0001733138196868822
step: 20, loss: 4.779957089340314e-05
step: 30, loss: 9.154379222309217e-05
step: 40, loss: 0.00013496940664481372
step: 50, loss: 0.00017490700702182949
step: 60, loss: 3.200293213012628e-05
step: 70, loss: 6.164383376017213e-05
step: 80, loss: 0.0014389686984941363
step: 90, loss: 2.7648333343677223e-05
step: 100, loss: 4.2969564674422145e-05
step: 110, loss: 4.356092540547252e-05
step: 120, loss: 8.961360435932875e-05
step: 130, loss: 6.134524301160127e-05
step: 140, loss: 0.0024714749306440353
step: 150, loss: 9.009404311655089e-05
step: 160, loss: 6.702243263134733e-05
step: 170, loss: 3.7832127418369055e-05
step: 180, loss: 0.00014388332783710212
step: 190, loss: 0.0002648008521646261
step: 200, loss: 5.2112951379967853e-05
step: 210, loss: 0.016961921006441116
step: 220, loss: 0.001429109601303935
step: 230, loss: 5.132795558893122e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9663677130044843, f1=0.967670011148272,
best_f1=0.9665924276169264
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True

```

```

loss_scale          : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2473it [00:10, 227.41it/s]
load_f1 = 0.9686800894854586
real_f1 = 0.9686800894854586
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:18, 236.75it/s]

```

3.3.2 DBLP-GoogleScholar - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
    --task Dirty/DBLP-GoogleScholar \
    --batch_size 32 \
    --max_len 256 \
    --lr 3e-5 \
    --n_epochs 15 \
    --finetuning \
    --lm distilbert \
    --fp16 \
    --save_model \
    --dk general \
    --dk product \
    --summarize

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-GoogleScholar/.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
    --task Dirty/DBLP-GoogleScholar \
    --input_path input/input_small.jsonl \

```

```

--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:

UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order: `optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

```

"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)

```

```
step: 0, loss: 0.791255533695221
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0

```
step: 10, loss: 0.4486081004142761
```

```
step: 20, loss: 0.5038030743598938
```

```
step: 30, loss: 0.437298059463501
```

step: 40, loss: 0.3761184811592102
step: 50, loss: 0.19969919323921204
step: 60, loss: 0.214988112449646
step: 70, loss: 0.19883033633232117
step: 80, loss: 0.16295276582241058
step: 90, loss: 0.09316208958625793
step: 100, loss: 0.32215777039527893
step: 110, loss: 0.07137736678123474
step: 120, loss: 0.02812262438237667
step: 130, loss: 0.023407313972711563
step: 140, loss: 0.13481269776821136
step: 150, loss: 0.02560950256884098
step: 160, loss: 0.23763561248779297
step: 170, loss: 0.11771588027477264
step: 180, loss: 0.17302827537059784
step: 190, loss: 0.048616159707307816
step: 200, loss: 0.07367201894521713
step: 210, loss: 0.12819597125053406
step: 220, loss: 0.20266282558441162
step: 230, loss: 0.07803527265787125
step: 240, loss: 0.12422942370176315
step: 250, loss: 0.056961044669151306
step: 260, loss: 0.11235906928777695
step: 270, loss: 0.01811787113547325
step: 280, loss: 0.07777007669210434
step: 290, loss: 0.15218903124332428
step: 300, loss: 0.05066283419728279
step: 310, loss: 0.07583274692296982
step: 320, loss: 0.0526898019015789
step: 330, loss: 0.07800900191068649
step: 340, loss: 0.15934181213378906
step: 350, loss: 0.08891227841377258
step: 360, loss: 0.06658405810594559
step: 370, loss: 0.11278610676527023
step: 380, loss: 0.20491084456443787
step: 390, loss: 0.014537390321493149
step: 400, loss: 0.020452504977583885
step: 410, loss: 0.018229136243462563
step: 420, loss: 0.0037052747793495655
step: 430, loss: 0.06372405588626862
step: 440, loss: 0.05392566695809364
step: 450, loss: 0.04313359037041664
step: 460, loss: 0.19057324528694153
step: 470, loss: 0.19058196246623993
step: 480, loss: 0.30320531129837036
step: 490, loss: 0.04636039584875107
step: 500, loss: 0.012113654986023903
step: 510, loss: 0.08572914451360703

```

step: 520, loss: 0.052384477108716965
step: 530, loss: 0.0771690383553505
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9319664492078286, f1=0.9286047596826879,
best_f1=0.9286047596826879
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1444178819656372
step: 10, loss: 0.13417565822601318
step: 20, loss: 0.0629497841000557
step: 30, loss: 0.022540679201483727
step: 40, loss: 0.016242656856775284
step: 50, loss: 0.08556653559207916
step: 60, loss: 0.08430030941963196
step: 70, loss: 0.1840263158082962
step: 80, loss: 0.01691245660185814
step: 90, loss: 0.008049361407756805
step: 100, loss: 0.576839029788971
step: 110, loss: 0.02926945686340332
step: 120, loss: 0.0836315006017685
step: 130, loss: 0.02507946267724037
step: 140, loss: 0.0371004194021225
step: 150, loss: 0.059951651841402054
step: 160, loss: 0.017718549817800522
step: 170, loss: 0.15390440821647644
step: 180, loss: 0.006704364903271198
step: 190, loss: 0.029365500435233116
step: 200, loss: 0.04305412992835045
step: 210, loss: 0.008051321841776371
step: 220, loss: 0.16407223045825958
step: 230, loss: 0.008402853272855282
step: 240, loss: 0.12682703137397766
step: 250, loss: 0.01636955514550209
step: 260, loss: 0.02482825145125389
step: 270, loss: 0.15383651852607727
step: 280, loss: 0.08839331567287445
step: 290, loss: 0.0980403944849968
step: 300, loss: 0.03554829582571983
step: 310, loss: 0.06866011023521423
step: 320, loss: 0.16458505392074585
step: 330, loss: 0.021746182814240456

```



```

step: 340, loss: 0.00276503199711442
step: 350, loss: 0.10417815297842026
step: 360, loss: 0.006040667649358511
step: 370, loss: 0.005450280383229256
step: 380, loss: 0.0908181220293045
step: 390, loss: 0.064237080514431
step: 400, loss: 0.03798220306634903
step: 410, loss: 0.0004972416791133583
step: 420, loss: 0.05810258910059929
step: 430, loss: 0.03536553680896759
step: 440, loss: 0.011990852653980255
step: 450, loss: 0.026258155703544617
step: 460, loss: 0.206917867064476
step: 470, loss: 0.03626338019967079
step: 480, loss: 0.19078877568244934
step: 490, loss: 0.03292703628540039
step: 500, loss: 0.043512582778930664
step: 510, loss: 0.07556498795747757
step: 520, loss: 0.10121607035398483
step: 530, loss: 0.10138267278671265
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9394785847299814, f1=0.934752429430819,
best_f1=0.934752429430819
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0090132225304842
step: 10, loss: 0.04565225541591644
step: 20, loss: 0.1837516725063324
step: 30, loss: 0.21182525157928467
step: 40, loss: 0.0024513695389032364
step: 50, loss: 0.014396089129149914
step: 60, loss: 0.007563100196421146
step: 70, loss: 0.02645226940512657
step: 80, loss: 0.0014604111202061176
step: 90, loss: 0.05894727259874344
step: 100, loss: 0.012508595362305641
step: 110, loss: 0.03345462679862976
step: 120, loss: 0.0033610414247959852
step: 130, loss: 0.03806142508983612
step: 140, loss: 0.053816333413124084
step: 150, loss: 0.005014392547309399

```

```

step: 160, loss: 0.004124544560909271
step: 170, loss: 0.003674211213365197
step: 180, loss: 0.011353657580912113
step: 190, loss: 0.011991789564490318
step: 200, loss: 0.01075421366840601
step: 210, loss: 0.031003080308437347
step: 220, loss: 0.009415464475750923
step: 230, loss: 0.004868949763476849
step: 240, loss: 0.01868022233247757
step: 250, loss: 0.004504588432610035
step: 260, loss: 0.006225323770195246
step: 270, loss: 0.0034021656028926373
step: 280, loss: 0.014237334951758385
step: 290, loss: 0.053976546972990036
step: 300, loss: 0.16202548146247864
step: 310, loss: 0.08014634996652603
step: 320, loss: 0.10541824251413345
step: 330, loss: 0.0072192782536149025
step: 340, loss: 0.001402436406351626
step: 350, loss: 0.10897226631641388
step: 360, loss: 0.03995222970843315
step: 370, loss: 0.0025858960580080748
step: 380, loss: 0.04195591062307358
step: 390, loss: 0.015175128355622292
step: 400, loss: 0.010108222253620625
step: 410, loss: 0.006328606512397528
step: 420, loss: 0.07529684156179428
step: 430, loss: 0.07550115883350372
step: 440, loss: 0.052090123295784
step: 450, loss: 0.05756634846329689
step: 460, loss: 0.062061961740255356
step: 470, loss: 0.005904640536755323
step: 480, loss: 0.0037200080696493387
step: 490, loss: 0.0032787774689495564
step: 500, loss: 0.013974908739328384
step: 510, loss: 0.007817672565579414
step: 520, loss: 0.002141684526577592
step: 530, loss: 0.03186500817537308
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.932666060054595, f1=0.9285067873303169,
best_f1=0.934752429430819
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06528593599796295
step: 10, loss: 0.016139285638928413

```

step: 20, loss: 0.019930917769670486
step: 30, loss: 0.0008622086024843156
step: 40, loss: 0.001837379764765501
step: 50, loss: 0.011662263423204422
step: 60, loss: 0.00043758717947639525
step: 70, loss: 0.004958657547831535
step: 80, loss: 0.006807961966842413
step: 90, loss: 0.0005554844974540174
step: 100, loss: 0.010760453529655933
step: 110, loss: 0.0002735452726483345
step: 120, loss: 0.004854696337133646
step: 130, loss: 0.21245872974395752
step: 140, loss: 0.07083530724048615
step: 150, loss: 0.032341230660676956
step: 160, loss: 0.00968270655721426
step: 170, loss: 0.052039697766304016
step: 180, loss: 0.010483231395483017
step: 190, loss: 0.029609397053718567
step: 200, loss: 0.0006342437118291855
step: 210, loss: 0.006222896743565798
step: 220, loss: 0.001423210953362286
step: 230, loss: 0.1843787431716919
step: 240, loss: 0.006432629656046629
step: 250, loss: 0.08220455050468445
step: 260, loss: 0.08724097907543182
step: 270, loss: 0.0030032345093786716
step: 280, loss: 0.0018593673594295979
step: 290, loss: 0.012818773277103901
step: 300, loss: 0.00045782618690282106
step: 310, loss: 0.0007015369483269751
step: 320, loss: 0.0019080983474850655
step: 330, loss: 0.015515148639678955
step: 340, loss: 0.007249138318002224
step: 350, loss: 0.0010345649207010865
step: 360, loss: 0.04289902374148369
step: 370, loss: 0.01162468921393156
step: 380, loss: 0.005346055142581463
step: 390, loss: 0.08158654719591141
step: 400, loss: 0.0049245040863752365
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 410, loss: 0.004459989257156849
step: 420, loss: 0.0028258359525352716
step: 430, loss: 0.027644654735922813
step: 440, loss: 0.07783593982458115
step: 450, loss: 0.011635278351604939
step: 460, loss: 0.00292919110506773
step: 470, loss: 0.008407428860664368
step: 480, loss: 0.0015127214137464762

```

step: 490, loss: 0.025365909561514854
step: 500, loss: 0.005806750152260065
step: 510, loss: 0.0040560513734817505
step: 520, loss: 0.003025258891284466
step: 530, loss: 0.010112659074366093
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.9420560747663551, f1=0.935831381733021,
best_f1=0.935831381733021
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0038391989655792713
step: 10, loss: 0.002141081029549241
step: 20, loss: 0.0013164662523195148
step: 30, loss: 0.0006658960483036935
step: 40, loss: 0.03991144895553589
step: 50, loss: 0.031797874718904495
step: 60, loss: 0.0016558498609811068
step: 70, loss: 0.00021979236043989658
step: 80, loss: 0.00026074444758705795
step: 90, loss: 0.0003115706203971058
step: 100, loss: 0.0003976173175033182
step: 110, loss: 0.02458997815847397
step: 120, loss: 0.004709675908088684
step: 130, loss: 0.0005051987827755511
step: 140, loss: 0.0034217031206935644
step: 150, loss: 0.00029641378205269575
step: 160, loss: 0.10560325533151627
step: 170, loss: 0.0026985169388353825
step: 180, loss: 0.0008168633794412017
step: 190, loss: 0.0015393721405416727
step: 200, loss: 0.0006091479444876313
step: 210, loss: 0.000641800812445581
step: 220, loss: 0.0010673445649445057
step: 230, loss: 0.00035146326990798116
step: 240, loss: 0.04176614060997963
step: 250, loss: 0.00220372062176466
step: 260, loss: 0.017783720046281815
step: 270, loss: 0.07528012990951538
step: 280, loss: 0.029990682378411293
step: 290, loss: 0.001507519744336605
step: 300, loss: 0.06490249931812286

```

```

step: 310, loss: 0.0836915671825409
step: 320, loss: 0.005084997974336147
step: 330, loss: 0.001714935409836471
step: 340, loss: 0.001996798673644662
step: 350, loss: 0.00972450990229845
step: 360, loss: 0.014549940824508667
step: 370, loss: 0.004275730345398188
step: 380, loss: 0.008246461860835552
step: 390, loss: 0.0004735137918032706
step: 400, loss: 0.0052504995837807655
step: 410, loss: 0.00020379693887662143
step: 420, loss: 0.001402965746819973
step: 430, loss: 0.001253708265721798
step: 440, loss: 0.0011812008451670408
step: 450, loss: 0.005634070839732885
step: 460, loss: 0.011835365556180477
step: 470, loss: 0.027782108634710312
step: 480, loss: 0.005136764142662287
step: 490, loss: 0.0005037934170104563
step: 500, loss: 0.012519629672169685
step: 510, loss: 0.0010803494369611144
step: 520, loss: 0.0003062007308471948
step: 530, loss: 0.04357392340898514
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.9420491423273065, f1=0.9304063521718824,
best_f1=0.935831381733021
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06523985415697098
step: 10, loss: 0.0006866800249554217
step: 20, loss: 0.0002474735083524138
step: 30, loss: 0.00028803336317650974
step: 40, loss: 0.00042545367614366114
step: 50, loss: 0.0026708098594099283
step: 60, loss: 0.0001949536381289363
step: 70, loss: 0.0132125373929739
step: 80, loss: 0.0004227476892992854
step: 90, loss: 0.0006066537462174892
step: 100, loss: 0.007621814962476492
step: 110, loss: 0.0014936259249225259
step: 120, loss: 0.0002292411809321493
step: 130, loss: 0.0002528316399548203
step: 140, loss: 0.0036142540629953146
step: 150, loss: 0.00030999889713712037
step: 160, loss: 0.0025465022772550583

```

```

step: 170, loss: 0.0016023981152102351
step: 180, loss: 8.305234223371372e-05
step: 190, loss: 0.0008608694770373404
step: 200, loss: 0.0002884758869186044
step: 210, loss: 0.00028070894768461585
step: 220, loss: 0.00021485604520421475
step: 230, loss: 0.00012193894508527592
step: 240, loss: 0.0005802339292131364
step: 250, loss: 0.00013812145334668458
step: 260, loss: 0.00014556568930856884
step: 270, loss: 0.00011564130545593798
step: 280, loss: 0.0005788418930023909
step: 290, loss: 0.01725633442401886
step: 300, loss: 0.009639348834753036
step: 310, loss: 0.0001836695009842515
step: 320, loss: 0.0876917839050293
step: 330, loss: 0.0033482711296528578
step: 340, loss: 0.02393742837011814
step: 350, loss: 0.007038157898932695
step: 360, loss: 7.109388388926163e-05
step: 370, loss: 0.0002807473938446492
step: 380, loss: 0.028358235955238342
step: 390, loss: 0.10335693508386612
step: 400, loss: 0.0016085872193798423
step: 410, loss: 7.868563989177346e-05
step: 420, loss: 0.013227268122136593
step: 430, loss: 0.0002908235474023968
step: 440, loss: 0.0015147279482334852
step: 450, loss: 0.00042895093793049455
step: 460, loss: 0.003587595419958234
step: 470, loss: 0.00459324661642313
step: 480, loss: 0.02413114719092846
step: 490, loss: 0.00014850783918518573
step: 500, loss: 0.038963668048381805
step: 510, loss: 0.0002990509383380413
step: 520, loss: 0.0027824128046631813
step: 530, loss: 0.08003582805395126
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.9246901811248809, f1=0.9148325358851676,
best_f1=0.935831381733021
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000603702908847481
step: 10, loss: 0.004642369691282511
step: 20, loss: 0.03334880992770195

```

step: 30, loss: 0.00019807297212537378
step: 40, loss: 0.0006234010797925293
step: 50, loss: 0.002986317267641425
step: 60, loss: 0.13773375749588013
step: 70, loss: 0.21315771341323853
step: 80, loss: 0.0056563797406852245
step: 90, loss: 0.002166938502341509
step: 100, loss: 0.0017934886272996664
step: 110, loss: 0.00024733791360631585
step: 120, loss: 0.00032198213739320636
step: 130, loss: 0.0005183913744986057
step: 140, loss: 0.0020967053715139627
step: 150, loss: 8.899426757125184e-05
step: 160, loss: 0.0011775443563237786
step: 170, loss: 0.008341878652572632
step: 180, loss: 0.00038693397073075175
step: 190, loss: 9.18930018087849e-05
step: 200, loss: 0.0005964299780316651
step: 210, loss: 0.00012319948291406035
step: 220, loss: 0.001361711649224162
step: 230, loss: 0.017266251146793365
step: 240, loss: 0.005103851202875376
step: 250, loss: 0.0012528778752312064
step: 260, loss: 0.0002784823882393539
step: 270, loss: 6.637940532527864e-05
step: 280, loss: 0.07015388458967209
step: 290, loss: 0.02121657319366932
step: 300, loss: 0.012188044376671314
step: 310, loss: 0.00010132369789062068
step: 320, loss: 0.0027481364086270332
step: 330, loss: 0.00015846538008190691
step: 340, loss: 0.007913067005574703
step: 350, loss: 0.0024508482310920954
step: 360, loss: 0.0016609558369964361
step: 370, loss: 0.06275510787963867
step: 380, loss: 0.001084569958038628
step: 390, loss: 0.0005286267842166126
step: 400, loss: 0.00011535597150214016
step: 410, loss: 0.00031193799804896116
step: 420, loss: 0.00018334697233512998
step: 430, loss: 0.0002720980846788734
step: 440, loss: 0.00010009785182774067
step: 450, loss: 0.00014268883387558162
step: 460, loss: 0.0016992359887808561
step: 470, loss: 0.012668060138821602
step: 480, loss: 0.00013904650404583663
step: 490, loss: 0.03340182453393936
step: 500, loss: 0.00014549557818099856

```

step: 510, loss: 0.001150237862020731
step: 520, loss: 0.00018471878138370812
step: 530, loss: 0.00026405518292449415
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9342723004694835, f1=0.9253308128544424,
best_f1=0.935831381733021
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00112982839345932
step: 10, loss: 0.0007903522346168756
step: 20, loss: 0.0740932822227478
step: 30, loss: 0.0178332831710577
step: 40, loss: 0.0044807433150708675
step: 50, loss: 0.0016717161051928997
step: 60, loss: 0.00014549781917594373
step: 70, loss: 0.0005285820225253701
step: 80, loss: 0.0001345844502793625
step: 90, loss: 0.00025321950670331717
step: 100, loss: 0.00018732900207396597
step: 110, loss: 0.0005894431960768998
step: 120, loss: 0.0004103248647879809
step: 130, loss: 0.00037416693521663547
step: 140, loss: 3.406274117878638e-05
step: 150, loss: 0.0002664308703970164
step: 160, loss: 0.00011101832205895334
step: 170, loss: 0.00012870885257143527
step: 180, loss: 0.00020926048455294222
step: 190, loss: 8.825182885630056e-05
step: 200, loss: 0.00010352390381740406
step: 210, loss: 0.06226871535181999
step: 220, loss: 0.000373618007870391
step: 230, loss: 0.0005160685977898538
step: 240, loss: 0.00012340612011030316
step: 250, loss: 0.00028670934261754155
step: 260, loss: 0.0006757760420441628
step: 270, loss: 3.127225500065833e-05
step: 280, loss: 0.0084900027140975
step: 290, loss: 0.00029051920864731073
step: 300, loss: 4.765516860061325e-05
step: 310, loss: 0.019916735589504242
step: 320, loss: 7.80703776399605e-05
step: 330, loss: 0.03855808451771736
step: 340, loss: 0.08238240331411362
step: 350, loss: 0.08858301490545273
step: 360, loss: 0.0007418401073664427

```



```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 370, loss: 0.0003769329050555825
step: 380, loss: 0.016376208513975143
step: 390, loss: 0.0011588463094085455
step: 400, loss: 0.153720885515213
step: 410, loss: 0.0017917151562869549
step: 420, loss: 7.223174179671332e-05
step: 430, loss: 0.0026969911996275187
step: 440, loss: 0.0001924228563439101
step: 450, loss: 0.0001422456552973017
step: 460, loss: 0.00020121218403801322
step: 470, loss: 0.0003258071665186435
step: 480, loss: 4.518874266068451e-05
step: 490, loss: 0.0002873954363167286
step: 500, loss: 0.00031713731004856527
step: 510, loss: 0.00010875251609832048
step: 520, loss: 0.00019582420645747334
step: 530, loss: 0.00021188516984693706
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9411219286045434, f1=0.9319129226493746,
best_f1=0.935831381733021
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00023821853392291814
step: 10, loss: 0.0006178974872455001
step: 20, loss: 0.00040991429705172777
step: 30, loss: 0.0003006116021424532
step: 40, loss: 0.014479240402579308
step: 50, loss: 0.0002214884152635932
step: 60, loss: 0.00035573478089645505
step: 70, loss: 0.04905767738819122
step: 80, loss: 0.0007391920080408454
step: 90, loss: 0.029974350705742836
step: 100, loss: 0.0002733865403570235
step: 110, loss: 0.0013263903092592955
step: 120, loss: 9.592097921995446e-05
step: 130, loss: 4.785267083207145e-05
step: 140, loss: 0.002070816233754158
step: 150, loss: 0.00010948861745418981
step: 160, loss: 0.0006044735200703144
step: 170, loss: 5.133715967531316e-05
step: 180, loss: 0.00020561415294650942
step: 190, loss: 0.001276893075555563
step: 200, loss: 0.01211960893124342
step: 210, loss: 3.555635703378357e-05

```

```

step: 220, loss: 0.00015007442561909556
step: 230, loss: 0.0007694162777625024
step: 240, loss: 0.00021712144371122122
step: 250, loss: 7.891438144724816e-05
step: 260, loss: 0.0003939251764677465
step: 270, loss: 0.0005167193012312055
step: 280, loss: 8.62954548210837e-05
step: 290, loss: 3.9321650547208264e-05
step: 300, loss: 0.001530313165858388
step: 310, loss: 4.5272172428667545e-05
step: 320, loss: 5.252558185020462e-05
step: 330, loss: 0.0001437202881788835
step: 340, loss: 0.002290302189067006
step: 350, loss: 5.8198558690492064e-05
step: 360, loss: 0.007203247863799334
step: 370, loss: 0.0001609928294783458
step: 380, loss: 3.877478593494743e-05
step: 390, loss: 3.281494718976319e-05
step: 400, loss: 0.09013321995735168
step: 410, loss: 4.930051363771781e-05
step: 420, loss: 0.0007567509892396629
step: 430, loss: 9.642996883485466e-05
step: 440, loss: 5.702505586668849e-05
step: 450, loss: 2.8318796466919594e-05
step: 460, loss: 0.07195558398962021
step: 470, loss: 3.5369117540540174e-05
step: 480, loss: 0.00034915650030598044
step: 490, loss: 4.705009632743895e-05
step: 500, loss: 0.03069411963224411
step: 510, loss: 0.012943056412041187
step: 520, loss: 3.316831498523243e-05
step: 530, loss: 0.00010006132652051747
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9389846297158826, f1=0.9333954354913834,
best_f1=0.935831381733021
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.382393571082503e-05
step: 10, loss: 2.673583549039904e-05
step: 20, loss: 0.00040810409700497985
step: 30, loss: 3.158946492476389e-05
step: 40, loss: 4.715532486443408e-05
step: 50, loss: 0.0001868247054517269
step: 60, loss: 0.00027384949498809874
step: 70, loss: 0.0015773813938722014

```

```
step: 80, loss: 0.0009874369716271758
step: 90, loss: 0.00010282963194185868
step: 100, loss: 9.169125405605882e-05
step: 110, loss: 0.013239873573184013
step: 120, loss: 8.631267701275647e-05
step: 130, loss: 5.9346009948058054e-05
step: 140, loss: 0.00018783060659188777
step: 150, loss: 4.024085865239613e-05
step: 160, loss: 0.00016722081636544317
step: 170, loss: 0.08967789262533188
step: 180, loss: 0.0008024833514355123
step: 190, loss: 4.883817382506095e-05
step: 200, loss: 0.0003490317321848124
step: 210, loss: 0.035590119659900665
step: 220, loss: 3.652448140201159e-05
step: 230, loss: 6.200662755873054e-05
step: 240, loss: 2.785338438116014e-05
step: 250, loss: 0.00018147366063203663
step: 260, loss: 0.0040259771049022675
step: 270, loss: 5.447418516268954e-05
step: 280, loss: 4.2583822505548596e-05
step: 290, loss: 5.551032882067375e-05
step: 300, loss: 0.00016597728244960308
step: 310, loss: 7.645972800673917e-05
step: 320, loss: 4.5314201997825876e-05
step: 330, loss: 0.00019299954874441028
step: 340, loss: 5.140774737810716e-05
step: 350, loss: 0.0002864550915546715
step: 360, loss: 0.0004463669320102781
step: 370, loss: 0.00010414326243335381
step: 380, loss: 3.140297849313356e-05
step: 390, loss: 3.8405989471357316e-05
step: 400, loss: 4.712815280072391e-05
step: 410, loss: 0.008662770502269268
step: 420, loss: 7.451149576809257e-05
step: 430, loss: 3.655087493825704e-05
step: 440, loss: 2.6262760002282448e-05
step: 450, loss: 0.0005314418231137097
step: 460, loss: 0.00023139496624935418
step: 470, loss: 2.66276765614748e-05
step: 480, loss: 0.00030641129706054926
step: 490, loss: 0.0401458665728569
step: 500, loss: 0.029760297387838364
step: 510, loss: 0.0010135115589946508
step: 520, loss: 0.0014297999441623688
step: 530, loss: 5.564644015976228e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```

    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9418386491557222, f1=0.9253871421867668,
best_f1=0.935831381733021
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.36273718252778e-05
step: 10, loss: 0.0002863469999283552
step: 20, loss: 3.374585503479466e-05
step: 30, loss: 0.022457681596279144
step: 40, loss: 0.0002409022708889097
step: 50, loss: 0.00032793369609862566
step: 60, loss: 3.7596681067952886e-05
step: 70, loss: 0.0008567332406528294
step: 80, loss: 0.011187057942152023
step: 90, loss: 0.0001157677688752301
step: 100, loss: 3.197885598638095e-05
step: 110, loss: 4.010319025837816e-05
step: 120, loss: 7.765460759401321e-05
step: 130, loss: 6.566868250956759e-05
step: 140, loss: 2.2641566829406656e-05
step: 150, loss: 3.4140855859732255e-05
step: 160, loss: 0.002098730532452464
step: 170, loss: 0.0016298647969961166
step: 180, loss: 7.169973105192184e-05
step: 190, loss: 5.1169368816772476e-05
step: 200, loss: 0.0008229681407101452
step: 210, loss: 5.918397073401138e-05
step: 220, loss: 0.0001946294796653092
step: 230, loss: 0.00012972942204214633
step: 240, loss: 5.646795762004331e-05
step: 250, loss: 5.20853755006101e-05
step: 260, loss: 3.144991205772385e-05
step: 270, loss: 0.0006469262298196554
step: 280, loss: 3.8564667192986235e-05
step: 290, loss: 0.0002824499970301986
step: 300, loss: 0.0004104773106519133
step: 310, loss: 0.00014191980881150812
step: 320, loss: 0.01906191185116768
step: 330, loss: 0.0002570623473729938
step: 340, loss: 3.337295856908895e-05
step: 350, loss: 0.0008593134698458016
step: 360, loss: 8.726456144358963e-05
step: 370, loss: 2.0283776393625885e-05
step: 380, loss: 0.00010300448047928512
step: 390, loss: 8.872303442331031e-05
step: 400, loss: 2.6239540602546185e-05
step: 410, loss: 1.6819478332763538e-05

```

```

step: 420, loss: 3.898305294569582e-05
step: 430, loss: 0.001025163452140987
step: 440, loss: 2.827788921422325e-05
step: 450, loss: 0.0006181415519677103
step: 460, loss: 0.0007789157098159194
step: 470, loss: 0.00022972506121732295
step: 480, loss: 8.835402695694938e-05
step: 490, loss: 0.00014138835831545293
step: 500, loss: 4.334729601396248e-05
step: 510, loss: 3.1235336791723967e-05
step: 520, loss: 2.388596476521343e-05
step: 530, loss: 8.95428893272765e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9363117870722433, f1=0.9230038022813687,
best_f1=0.935831381733021
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.336047176271677e-05
step: 10, loss: 7.115586049621925e-05
step: 20, loss: 2.3845112082199194e-05
step: 30, loss: 0.0006937248399481177
step: 40, loss: 0.00011584367166506127
step: 50, loss: 0.00047041510697454214
step: 60, loss: 0.0009166153613477945
step: 70, loss: 0.00013520616630557925
step: 80, loss: 0.002682811114937067
step: 90, loss: 9.357522503705695e-05
step: 100, loss: 0.00010622133413562551
step: 110, loss: 6.706721615046263e-05
step: 120, loss: 6.407965702237561e-05
step: 130, loss: 3.94188282371033e-05
step: 140, loss: 4.398350938572548e-05
step: 150, loss: 0.0007623510318808258
step: 160, loss: 4.4796492147725075e-05
step: 170, loss: 0.00029947093571536243
step: 180, loss: 0.0017002192325890064
step: 190, loss: 6.315061182249337e-05
step: 200, loss: 7.366079807979986e-05
step: 210, loss: 0.0003629429265856743
step: 220, loss: 2.6568011890049092e-05
step: 230, loss: 0.0005545421154238284
step: 240, loss: 4.8482510464964435e-05
step: 250, loss: 2.8054177164449356e-05
step: 260, loss: 0.0002070825721602887
step: 270, loss: 0.001519912388175726

```

```

step: 280, loss: 3.106667281826958e-05
step: 290, loss: 0.006067801732569933
step: 300, loss: 0.00013048038817942142
step: 310, loss: 8.179806900443509e-05
step: 320, loss: 3.2810909033287317e-05
step: 330, loss: 2.6773008357849903e-05
step: 340, loss: 0.12620273232460022
step: 350, loss: 0.003408085322007537
step: 360, loss: 0.0003839692799374461
step: 370, loss: 2.954059891635552e-05
step: 380, loss: 7.376033317996189e-05
step: 390, loss: 4.7785033530090004e-05
step: 400, loss: 0.0009278040961362422
step: 410, loss: 0.013885554857552052
step: 420, loss: 0.0007013258873485029
step: 430, loss: 5.657813380821608e-05
step: 440, loss: 5.97562211623881e-05
step: 450, loss: 4.738452480523847e-05
step: 460, loss: 0.0037287401501089334
step: 470, loss: 0.021321704611182213
step: 480, loss: 6.302444671746343e-05
step: 490, loss: 9.577611490385607e-05
step: 500, loss: 0.0006038938299752772
step: 510, loss: 0.00023602554574608803
step: 520, loss: 0.005792319308966398
step: 530, loss: 0.00010325726179871708
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9391143911439114, f1=0.9366034243405831,
best_f1=0.935831381733021
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.013704029843211174
step: 10, loss: 0.00010584897972876206
step: 20, loss: 0.01865711249411106
step: 30, loss: 0.0004426601226441562
step: 40, loss: 6.322383705992252e-05
step: 50, loss: 7.231334166135639e-05
step: 60, loss: 7.970616570673883e-05
step: 70, loss: 0.0001127090654335916
step: 80, loss: 2.3338390747085214e-05
step: 90, loss: 0.0006966827786527574
step: 100, loss: 0.00012376382073853165
step: 110, loss: 0.00036253250436857343
step: 120, loss: 7.652064232388511e-05

```

```

step: 130, loss: 0.00018655764870345592
step: 140, loss: 9.528539521852508e-05
step: 150, loss: 4.609579991665669e-05
step: 160, loss: 2.9138300305930898e-05
step: 170, loss: 8.51929362397641e-05
step: 180, loss: 3.18644852086436e-05
step: 190, loss: 7.78366593294777e-05
step: 200, loss: 0.0023559220135211945
step: 210, loss: 0.00017450618906877935
step: 220, loss: 0.00044164914288558066
step: 230, loss: 2.676899930520449e-05
step: 240, loss: 4.848800745094195e-05
step: 250, loss: 0.04854520782828331
step: 260, loss: 2.5752260626177303e-05
step: 270, loss: 0.00013764435425400734
step: 280, loss: 0.0008976996759884059
step: 290, loss: 4.1588918975321576e-05
step: 300, loss: 5.347657497623004e-05
step: 310, loss: 2.283889989485033e-05
step: 320, loss: 8.706208609510213e-05
step: 330, loss: 0.0011263636406511068
step: 340, loss: 3.056433706660755e-05
step: 350, loss: 0.017860563471913338
step: 360, loss: 2.4850143745425157e-05
step: 370, loss: 4.30669570050668e-05
step: 380, loss: 0.00018306677520740777
step: 390, loss: 0.0003660774091258645
step: 400, loss: 7.33182969270274e-05
step: 410, loss: 0.0003747942973859608
step: 420, loss: 3.376718086656183e-05
step: 430, loss: 0.004962078761309385
step: 440, loss: 5.474662611959502e-05
step: 450, loss: 4.052861550007947e-05
step: 460, loss: 2.6392015570309013e-05
step: 470, loss: 0.00666930852457881
step: 480, loss: 0.00017903937259688973
step: 490, loss: 3.4673015761654824e-05
step: 500, loss: 0.00027127869543619454
step: 510, loss: 3.807733810390346e-05
step: 520, loss: 3.982537964475341e-05
step: 530, loss: 5.670985046890564e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)

```

```
epoch 13: dev_f1=0.9428838951310862, f1=0.9325210871602625,
best_f1=0.9325210871602625
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.843089143629186e-05
step: 10, loss: 6.41348451608792e-05
step: 20, loss: 1.6841833712533116e-05
step: 30, loss: 3.056512650800869e-05
step: 40, loss: 0.00010460240446263924
step: 50, loss: 0.00012847327161580324
step: 60, loss: 2.313329969183542e-05
step: 70, loss: 3.363935320521705e-05
step: 80, loss: 0.00013931805733591318
step: 90, loss: 1.6040887203416787e-05
step: 100, loss: 4.251649079378694e-05
step: 110, loss: 2.818070788634941e-05
step: 120, loss: 2.1225983800832182e-05
step: 130, loss: 0.0006694893818348646
step: 140, loss: 0.00013998393842484802
step: 150, loss: 9.680366929387674e-05
step: 160, loss: 0.001179753104224801
step: 170, loss: 4.2674328142311424e-05
step: 180, loss: 0.00017263967311009765
step: 190, loss: 0.00010438381286803633
step: 200, loss: 8.369411807507277e-05
step: 210, loss: 0.004462176468223333
step: 220, loss: 2.1948979338048957e-05
step: 230, loss: 0.0002594238903839141
step: 240, loss: 4.169384919805452e-05
step: 250, loss: 3.213289164705202e-05
step: 260, loss: 1.658845394558739e-05
step: 270, loss: 0.0006959849270060658
step: 280, loss: 2.1181334886932746e-05
step: 290, loss: 5.170809163246304e-05
step: 300, loss: 2.6247034838888794e-05
step: 310, loss: 2.345005668757949e-05
step: 320, loss: 0.00040576187893748283
step: 330, loss: 0.10581496357917786
step: 340, loss: 4.743485988001339e-05
step: 350, loss: 0.0008185053593479097
step: 360, loss: 0.001235149335116148
step: 370, loss: 5.3787578508490697e-05
step: 380, loss: 1.5396451999549754e-05
step: 390, loss: 0.0006923277396708727
step: 400, loss: 4.552928658085875e-05
step: 410, loss: 2.4414923245785758e-05
step: 420, loss: 1.955746665771585e-05
```



```

step: 430, loss: 2.26156826101942e-05
step: 440, loss: 1.448374860046897e-05
step: 450, loss: 2.8448977900552563e-05
step: 460, loss: 0.00031774266972206533
step: 470, loss: 2.0112351194256917e-05
step: 480, loss: 2.559942367952317e-05
step: 490, loss: 2.3967928427737206e-05
step: 500, loss: 0.00026986096054315567
step: 510, loss: 2.2641841496806592e-05
step: 520, loss: 8.031161269173026e-05
step: 530, loss: 2.2932463252800517e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9418931583880038, f1=0.9303857008466604,
best_f1=0.9325210871602625
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.290187204489484e-05
step: 10, loss: 4.725091275759041e-05
step: 20, loss: 2.7212437998969108e-05
step: 30, loss: 3.2147621823241934e-05
step: 40, loss: 3.334493885631673e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 50, loss: 3.505793210933916e-05
step: 60, loss: 5.4375042964238673e-05
step: 70, loss: 0.00017139325791504234
step: 80, loss: 3.66795138688758e-05
step: 90, loss: 6.991673581069335e-05
step: 100, loss: 3.337175803608261e-05
step: 110, loss: 0.000514268409460783
step: 120, loss: 0.00017521949484944344
step: 130, loss: 1.8663167793420143e-05
step: 140, loss: 1.6312802472384647e-05
step: 150, loss: 0.04034152254462242
step: 160, loss: 4.21922595705837e-05
step: 170, loss: 3.753774944925681e-05
step: 180, loss: 3.199215279892087e-05
step: 190, loss: 8.401969535043463e-05
step: 200, loss: 3.594564623199403e-05
step: 210, loss: 2.2943419025978073e-05
step: 220, loss: 1.7299938917858526e-05
step: 230, loss: 1.9430761312833056e-05
step: 240, loss: 3.255812043789774e-05
step: 250, loss: 0.00021279272914398462
step: 260, loss: 3.460951484157704e-05
step: 270, loss: 7.832355186110362e-05

```

```

step: 280, loss: 7.491173164453357e-05
step: 290, loss: 0.0006567100062966347
step: 300, loss: 0.00029016556800343096
step: 310, loss: 0.00017146149184554815
step: 320, loss: 0.00017186968761961907
step: 330, loss: 6.266549462452531e-05
step: 340, loss: 0.0001788273366400972
step: 350, loss: 0.0001457434264011681
step: 360, loss: 0.000917366414796561
step: 370, loss: 1.9196077118976973e-05
step: 380, loss: 3.2801443012431264e-05
step: 390, loss: 2.590424264781177e-05
step: 400, loss: 5.089013939141296e-05
step: 410, loss: 0.00012778578093275428
step: 420, loss: 2.9283357434906065e-05
step: 430, loss: 3.315689536975697e-05
step: 440, loss: 8.031679317355156e-05
step: 450, loss: 0.0001126863862737082
step: 460, loss: 1.5593865100527182e-05
step: 470, loss: 2.6351028282078914e-05
step: 480, loss: 0.00014294954598881304
step: 490, loss: 6.051745003787801e-05
step: 500, loss: 0.0003281783137936145
step: 510, loss: 0.00010525604739086702
step: 520, loss: 4.84533047711011e-05
step: 530, loss: 5.504591536009684e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9423697556477639, f1=0.9367789570835257,
best_f1=0.9325210871602625
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16

```

```

patch_torch_functions : False
keep_batchnorm_fp32   : True
master_weights        : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
5742it [00:23, 243.56it/s]
load_f1 = 0.9355294117647059
real_f1 = 0.9338338808071328
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:17, 255.49it/s]

```

4 DITTO

4.1 DITTO STRUCTURED

4.1.1 Beer - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
--task Structured/Beer \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \
--dk product \
--summarize \
--da entry_swap \

```

```

--da attr_del

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Beer.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/Beer \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:

UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`. In PyTorch 1.1.0 and later, you should call them in the opposite order:

`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more

details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

```

"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8549522757530212
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.2947368421052632, f1=0.3076923076923077,
best_f1=0.3076923076923077
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.37953177094459534
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.32098765432098764, f1=0.32876712328767127,
best_f1=0.32876712328767127
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3290550410747528
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.5, f1=0.3333333333333333, best_f1=0.3333333333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3602626621723175
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.3661971830985915, f1=0.38095238095238093,

```

```

best_f1=0.3333333333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3235913813114166
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.4545454545454545, f1=0.391304347826087,
best_f1=0.3333333333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3226291537284851
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.4878048780487805, f1=0.37209302325581395,
best_f1=0.3333333333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2527400553226471
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.5161290322580646, f1=0.23999999999999996,
best_f1=0.23999999999999996
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.32357755303382874
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 8: dev_f1=0.5625000000000001, f1=0.4444444444444444,
best_f1=0.4444444444444444
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.21267321705818176
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 9: dev_f1=0.5806451612903226, f1=0.4242424242424242,
best_f1=0.4242424242424242
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.28860294818878174
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 10: dev_f1=0.7200000000000001, f1=0.43478260869565216,
best_f1=0.43478260869565216
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.287803053855896
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.6666666666666666, f1=0.39999999999999997,
best_f1=0.43478260869565216
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2894260585308075
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.6666666666666666, f1=0.47058823529411764,
best_f1=0.43478260869565216
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13893042504787445
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

epoch 13: dev_f1=0.6666666666666666, f1=0.5789473684210527,
best_f1=0.43478260869565216
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2647566497325897
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.6666666666666666, f1=0.5405405405405405,
best_f1=0.43478260869565216
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.30301395058631897
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.6666666666666666, f1=0.5405405405405405,
best_f1=0.43478260869565216
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

```

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

91it [00:00, 133735.69it/s]

```



```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
load_f1 = 0.5925925925925927
real_f1 = 0.5925925925925927
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:17, 258.51it/s]

```

4.1.2 DBLP-ACM - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
--task Structured/DBLP-ACM \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \
--dk product \
--summarize \
--da entry_swap \
--da attr_del

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-ACM.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
--task Structured/DBLP-ACM \
--input_path input/input_small.jsonl \
--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:

UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`. In PyTorch 1.1.0 and later, you should call them in the opposite order: `optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

```
"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
```

```
step: 0, loss: 0.8077124357223511
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0

```
step: 10, loss: 0.46410495042800903
```

```
step: 20, loss: 0.602811872959137
```

```
step: 30, loss: 0.4405469000339508
```

```
step: 40, loss: 0.26537254452705383
```

```
step: 50, loss: 0.055750757455825806
```

```
step: 60, loss: 0.1531224548816681
```

```
step: 70, loss: 0.16512998938560486
```

```
step: 80, loss: 0.15678422152996063
```

```
step: 90, loss: 0.02232767455279827
```

```
step: 100, loss: 0.044891562312841415
```

```

step: 110, loss: 0.0180195402354002
step: 120, loss: 0.10375132411718369
step: 130, loss: 0.013753454200923443
step: 140, loss: 0.07544776797294617
step: 150, loss: 0.09456273168325424
step: 160, loss: 0.15723177790641785
step: 170, loss: 0.013527262955904007
step: 180, loss: 0.01062020193785429
step: 190, loss: 0.03410552442073822
step: 200, loss: 0.01315243449062109
step: 210, loss: 0.012562455609440804
step: 220, loss: 0.018344756215810776
step: 230, loss: 0.013798778876662254
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9720670391061451, f1=0.9707207207207207,
best_f1=0.9707207207207207
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004625513684004545
step: 10, loss: 0.003741889027878642
step: 20, loss: 0.0032720656599849463
step: 30, loss: 0.001862377393990755
step: 40, loss: 0.007716807071119547
step: 50, loss: 0.004130543675273657
step: 60, loss: 0.002885657362639904
step: 70, loss: 0.002754155546426773
step: 80, loss: 0.14699062705039978
step: 90, loss: 0.005284757819026709
step: 100, loss: 0.010265634395182133
step: 110, loss: 0.06406128406524658
step: 120, loss: 0.0017558385152369738
step: 130, loss: 0.0032946032006293535
step: 140, loss: 0.22361823916435242
step: 150, loss: 0.008044282905757427
step: 160, loss: 0.037736956030130386
step: 170, loss: 0.039295028895139694
step: 180, loss: 0.003581504337489605
step: 190, loss: 0.03850488364696503
step: 200, loss: 0.0012283582473173738
step: 210, loss: 0.03080006130039692
step: 220, loss: 0.0015643143560737371

```

```

step: 230, loss: 0.014551393687725067
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9843400447427293, f1=0.9809203142536477,
best_f1=0.9809203142536477
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06869376450777054
step: 10, loss: 0.015242524445056915
step: 20, loss: 0.009163403883576393
step: 30, loss: 0.03495381027460098
step: 40, loss: 0.06590628623962402
step: 50, loss: 0.002403519582003355
step: 60, loss: 0.05986183136701584
step: 70, loss: 0.00051729945698753
step: 80, loss: 0.0011126797180622816
step: 90, loss: 0.003357921727001667
step: 100, loss: 0.0005023599951528013
step: 110, loss: 0.007016051560640335
step: 120, loss: 0.001263291691429913
step: 130, loss: 0.000953886192291975
step: 140, loss: 0.023630226030945778
step: 150, loss: 0.0007362192845903337
step: 160, loss: 0.007216074503958225
step: 170, loss: 0.002337351441383362
step: 180, loss: 0.001147697912529111
step: 190, loss: 0.003384023206308484
step: 200, loss: 0.0010666897287592292
step: 210, loss: 0.0285679679363966
step: 220, loss: 0.007768651936203241
step: 230, loss: 0.08397001028060913
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.9887133182844244, f1=0.9807474518686297,
best_f1=0.9807474518686297
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0027829273603856564
step: 10, loss: 0.0006662656087428331
step: 20, loss: 0.0018525829073041677
step: 30, loss: 0.00039934294181875885
step: 40, loss: 0.000909722235519439
step: 50, loss: 0.0026417861226946115
step: 60, loss: 0.002567658433690667
step: 70, loss: 0.021647898480296135
step: 80, loss: 0.05547214671969414
step: 90, loss: 0.002174025634303689
step: 100, loss: 0.004859728738665581
step: 110, loss: 0.001747753587551415
step: 120, loss: 0.001879589050076902
step: 130, loss: 0.0034698175732046366
step: 140, loss: 0.05124927684664726
step: 150, loss: 0.027013253420591354
step: 160, loss: 0.004863815847784281
step: 170, loss: 0.012723945081233978
step: 180, loss: 0.03772178292274475
step: 190, loss: 0.011205406859517097
step: 200, loss: 0.0004685705353040248
step: 210, loss: 0.000548573094420135
step: 220, loss: 0.000496903492603451
step: 230, loss: 0.0002892251359298825
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.9854096520763187, f1=0.9864559819413092,
best_f1=0.9807474518686297
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014248797670006752
step: 10, loss: 0.00025170817389152944
step: 20, loss: 0.00029069979791529477
step: 30, loss: 0.00011639681179076433
step: 40, loss: 0.04375394433736801
step: 50, loss: 0.000427153951022774
step: 60, loss: 0.002347606234252453
step: 70, loss: 0.10622075945138931
step: 80, loss: 0.0017713992856442928
step: 90, loss: 0.005924344062805176
step: 100, loss: 0.01831200160086155
step: 110, loss: 0.0011465076822787523
step: 120, loss: 0.08335252106189728
step: 130, loss: 0.01563824899494648
step: 140, loss: 0.001732496079057455

```

```

step: 150, loss: 0.00020806555403396487
step: 160, loss: 0.007629320025444031
step: 170, loss: 0.07463763654232025
step: 180, loss: 0.002451870357617736
step: 190, loss: 0.0006247235578484833
step: 200, loss: 0.007319396827369928
step: 210, loss: 0.009960723109543324
step: 220, loss: 0.00025662517873570323
step: 230, loss: 0.0001853439025580883
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.983050847457627, f1=0.9759999999999999,
best_f1=0.9807474518686297
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009552503004670143
step: 10, loss: 0.0006404871819540858
step: 20, loss: 0.003016790607944131
step: 30, loss: 0.004473390057682991
step: 40, loss: 0.00041306717321276665
step: 50, loss: 0.0007546675042249262
step: 60, loss: 0.015349102206528187
step: 70, loss: 0.00020420474174898118
step: 80, loss: 0.0017831174191087484
step: 90, loss: 0.000614139367826283
step: 100, loss: 0.00021350114548113197
step: 110, loss: 0.06862108409404755
step: 120, loss: 0.0007635417860001326
step: 130, loss: 0.005375965032726526
step: 140, loss: 0.002294109435752034
step: 150, loss: 0.0007935070898383856
step: 160, loss: 0.10546518117189407
step: 170, loss: 0.0018328293226659298
step: 180, loss: 0.0012779423268511891
step: 190, loss: 0.0004314756370149553
step: 200, loss: 0.0061190929263830185
step: 210, loss: 0.0010553973261266947
step: 220, loss: 0.0008865909767337143
step: 230, loss: 0.0007337337010540068
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.9865470852017937, f1=0.9841628959276018,
best_f1=0.9807474518686297
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13372978568077087
step: 10, loss: 0.0002979643177241087
step: 20, loss: 0.00041702770977281034
step: 30, loss: 0.0004092722956556827
step: 40, loss: 0.016584137454628944
step: 50, loss: 0.0001376323343720287
step: 60, loss: 0.0002538668632041663
step: 70, loss: 0.0003450356307439506
step: 80, loss: 0.0002211074170190841
step: 90, loss: 0.000899057777132839
step: 100, loss: 0.0003220598155166954
step: 110, loss: 0.00930827297270298
step: 120, loss: 0.0010367684299126267
step: 130, loss: 0.003343655727803707
step: 140, loss: 0.0004428671963978559
step: 150, loss: 0.0025294460356235504
step: 160, loss: 0.0005316377501003444
step: 170, loss: 0.0011838204227387905
step: 180, loss: 0.023288363590836525
step: 190, loss: 0.0004550788435153663
step: 200, loss: 0.016796162351965904
step: 210, loss: 0.0002002450346481055
step: 220, loss: 0.004140470176935196
step: 230, loss: 0.03670358285307884
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9853438556933484, f1=0.9795454545454545,
best_f1=0.9807474518686297
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04665336757898331
step: 10, loss: 0.40363046526908875
step: 20, loss: 0.00029357464518398046
step: 30, loss: 0.0009290294838137925
step: 40, loss: 0.0008584341849200428
step: 50, loss: 0.007303754333406687
step: 60, loss: 0.001866282895207405
step: 70, loss: 0.0006941761239431798
step: 80, loss: 0.0008314821170642972
step: 90, loss: 0.00017067411681637168
step: 100, loss: 0.0002566428738646209
step: 110, loss: 0.0002801111259032041
step: 120, loss: 0.00039372954051941633
step: 130, loss: 0.0010870505357161164
step: 140, loss: 8.958339458331466e-05

```

```

step: 150, loss: 9.410580969415605e-05
step: 160, loss: 0.001542022917419672
step: 170, loss: 0.0001655833184486255
step: 180, loss: 0.012986048124730587
step: 190, loss: 0.0007787491194903851
step: 200, loss: 0.0053150951862335205
step: 210, loss: 0.0006617146427743137
step: 220, loss: 0.0044555580243468285
step: 230, loss: 0.030604684725403786
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.987598647125141, f1=0.9807037457434733,
best_f1=0.9807474518686297
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02439088001847267
step: 10, loss: 0.0011603852035477757
step: 20, loss: 0.00629135686904192
step: 30, loss: 0.0027480677235871553
step: 40, loss: 7.469786214642227e-05
step: 50, loss: 0.00027394943754188716
step: 60, loss: 0.00014894008927512914
step: 70, loss: 0.0008884213748387992
step: 80, loss: 0.04209970310330391
step: 90, loss: 0.0006120183388702571
step: 100, loss: 0.0008476361399516463
step: 110, loss: 0.0001899437338579446
step: 120, loss: 0.21667523682117462
step: 130, loss: 0.0003964205679949373
step: 140, loss: 0.013067744672298431
step: 150, loss: 0.00013879699690733105
step: 160, loss: 0.0012873122468590736
step: 170, loss: 0.00019120874640066177
step: 180, loss: 0.00022359489230439067
step: 190, loss: 0.00016445736400783062
step: 200, loss: 0.0002546126488596201
step: 210, loss: 0.0005270933615975082
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 220, loss: 0.05220663547515869
step: 230, loss: 0.007467035204172134
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9864864864864865, f1=0.9829738933030647,
best_f1=0.9807474518686297
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```



```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00020931502513121814
step: 10, loss: 9.729675366543233e-05
step: 20, loss: 0.00010498751362320036
step: 30, loss: 0.00015100104792509228
step: 40, loss: 6.533411215059459e-05
step: 50, loss: 0.00493826437741518
step: 60, loss: 0.0369204543530941
step: 70, loss: 0.0012847766047343612
step: 80, loss: 0.00015580632316414267
step: 90, loss: 0.024406328797340393
step: 100, loss: 0.00013663643039762974
step: 110, loss: 0.00011774095037253574
step: 120, loss: 0.014301060698926449
step: 130, loss: 0.0015677014598622918
step: 140, loss: 0.023728851228952408
step: 150, loss: 9.710752783576027e-05
step: 160, loss: 0.0002255684812553227
step: 170, loss: 0.00017236715939361602
step: 180, loss: 0.0005851485766470432
step: 190, loss: 0.00011992535291938111
step: 200, loss: 0.00011424686817917973
step: 210, loss: 8.994058589451015e-05
step: 220, loss: 0.0019306916510686278
step: 230, loss: 0.00040941181941889226
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9843400447427293, f1=0.9808342728297633,
best_f1=0.9807474518686297
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.442563557764515e-05
step: 10, loss: 8.09077246231027e-05
step: 20, loss: 0.0009772334014996886
step: 30, loss: 0.0004415281873662025
step: 40, loss: 0.0001760700106387958
step: 50, loss: 0.0001775137207005173
step: 60, loss: 8.539349801139906e-05
step: 70, loss: 0.0002363072126172483
step: 80, loss: 0.0013243077555671334
step: 90, loss: 0.0006520449533127248
step: 100, loss: 0.00014940359687898308
step: 110, loss: 0.00014315429143607616
step: 120, loss: 0.00039323914097622037
step: 130, loss: 6.87090796418488e-05

```

```

step: 140, loss: 0.007226657588034868
step: 150, loss: 6.03084554313682e-05
step: 160, loss: 0.00019125583639834076
step: 170, loss: 0.010930919088423252
step: 180, loss: 0.11176751554012299
step: 190, loss: 0.00044521878589875996
step: 200, loss: 0.0002240262256236747
step: 210, loss: 8.193960093194619e-05
step: 220, loss: 0.0007396735600195825
step: 230, loss: 0.02198212966322899
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9865470852017937, f1=0.9864559819413092,
best_f1=0.9807474518686297
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0016292372019961476
step: 10, loss: 0.008584576658904552
step: 20, loss: 5.722037167288363e-05
step: 30, loss: 0.001892730826511979
step: 40, loss: 4.83700969198253e-05
step: 50, loss: 4.641824489226565e-05
step: 60, loss: 0.010927030816674232
step: 70, loss: 0.00048451792099513113
step: 80, loss: 5.407041680882685e-05
step: 90, loss: 5.941346171312034e-05
step: 100, loss: 7.149884186219424e-05
step: 110, loss: 0.00018138934683520347
step: 120, loss: 0.008621389977633953
step: 130, loss: 0.00029188382904976606
step: 140, loss: 8.452680049231276e-05
step: 150, loss: 7.516034384025261e-05
step: 160, loss: 0.0346924364566803
step: 170, loss: 0.00010610563913360238
step: 180, loss: 0.00044926113332621753
step: 190, loss: 5.129992132424377e-05
step: 200, loss: 0.013555807992815971
step: 210, loss: 3.961342736147344e-05
step: 220, loss: 0.02802310325205326
step: 230, loss: 9.925005724653602e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9887133182844244, f1=0.9783352337514253,
best_f1=0.9807474518686297
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 4.6339468099176884e-05
step: 10, loss: 4.460722266230732e-05
step: 20, loss: 0.00020104917348362505
step: 30, loss: 0.027907148003578186
step: 40, loss: 0.00014525117876473814
step: 50, loss: 7.778610597597435e-05
step: 60, loss: 0.002309780102223158
step: 70, loss: 4.241440910845995e-05
step: 80, loss: 0.008519927971065044
step: 90, loss: 0.00044642837019637227
step: 100, loss: 8.795934991212562e-05
step: 110, loss: 0.00011157045082654804
step: 120, loss: 4.831723708775826e-05
step: 130, loss: 8.128515764838085e-05
step: 140, loss: 0.00016169702576007694
step: 150, loss: 0.02463308349251747
step: 160, loss: 3.852903682854958e-05
step: 170, loss: 4.2004656279459596e-05
step: 180, loss: 7.847768574720249e-05
step: 190, loss: 7.568422734038904e-05
step: 200, loss: 4.3419651774456725e-05
step: 210, loss: 5.757362669100985e-05
step: 220, loss: 2.896352634707e-05
step: 230, loss: 4.3639447540044785e-05
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

```
epoch 13: dev_f1=0.987598647125141, f1=0.9783352337514253,
best_f1=0.9807474518686297
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 3.331054176669568e-05
step: 10, loss: 2.8110302082495764e-05
step: 20, loss: 0.014717942103743553
step: 30, loss: 8.880948735168204e-05
step: 40, loss: 3.2800078770378605e-05
step: 50, loss: 5.9479465562617406e-05
step: 60, loss: 0.0002024240675382316
step: 70, loss: 0.000508532568346709
step: 80, loss: 6.077886064304039e-05
step: 90, loss: 6.901226879563183e-05
step: 100, loss: 0.017867770045995712
step: 110, loss: 0.00020382220100145787
step: 120, loss: 7.387758523691446e-05
step: 130, loss: 5.845410851179622e-05
```

```

step: 140, loss: 0.0011641128221526742
step: 150, loss: 0.0001089785509975627
step: 160, loss: 3.7373247323557734e-05
step: 170, loss: 3.6692668800242245e-05
step: 180, loss: 6.614199082832783e-05
step: 190, loss: 0.000977959018200636
step: 200, loss: 3.5892106097890064e-05
step: 210, loss: 0.0015447696205228567
step: 220, loss: 0.00010658328392310068
step: 230, loss: 2.7179112294106744e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 14: dev_f1=0.9887387387387387, f1=0.9806598407281, best_f1=0.9806598407281
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.43001330900006e-05
step: 10, loss: 4.932714728056453e-05
step: 20, loss: 3.568337342585437e-05
step: 30, loss: 0.000466939847683534
step: 40, loss: 4.5707212848355994e-05
step: 50, loss: 0.00014174089301377535
step: 60, loss: 3.62443424819503e-05
step: 70, loss: 0.00015867839101701975
step: 80, loss: 0.04060671105980873
step: 90, loss: 2.9060232918709517e-05
step: 100, loss: 3.9438422390958294e-05
step: 110, loss: 0.00010385369387222454
step: 120, loss: 0.007464008405804634
step: 130, loss: 0.0004032192227896303
step: 140, loss: 0.00018011234351433814
step: 150, loss: 0.0004055554745718837
step: 160, loss: 0.00030042111757211387
step: 170, loss: 2.9678498322027735e-05
step: 180, loss: 0.00023065162531565875
step: 190, loss: 0.01334638986736536
step: 200, loss: 6.455100810853764e-05
step: 210, loss: 0.00043369235936552286
step: 220, loss: 0.00015392554632853717
step: 230, loss: 0.00014997771359048784
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 15: dev_f1=0.9887640449438202, f1=0.9852774631936579,
best_f1=0.9852774631936579
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : O2
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
2473it [00:10, 233.20it/s]
load_f1 = 0.9854096520763187
real_f1 = 0.9854096520763187
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:17, 252.10it/s]

```

4.1.3 DBLP-GoogleScholar - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
--task Structured/DBLP-GoogleScholar \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \
--dk product \
--summarize \
--da entry_swap \
--da attr_del

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-GoogleScholar.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
--task Structured/DBLP-GoogleScholar \
--input_path input/input_small.jsonl \
--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

enabled	: True
opt_level	: 02
cast_model_type	: torch.float16
patch_torch_functions	: False
keep_batchnorm_fp32	: True
master_weights	: True
loss_scale	: dynamic

```

Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.7950355410575867
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4403623938560486
step: 20, loss: 0.4947464168071747
step: 30, loss: 0.4291040003299713
step: 40, loss: 0.34144002199172974
step: 50, loss: 0.21199959516525269
step: 60, loss: 0.10082711279392242
step: 70, loss: 0.12104038894176483
step: 80, loss: 0.15808963775634766
step: 90, loss: 0.2537749707698822
step: 100, loss: 0.23980966210365295
step: 110, loss: 0.05803566426038742
step: 120, loss: 0.0799383595585823
step: 130, loss: 0.04334482178092003
step: 140, loss: 0.1285054087638855
step: 150, loss: 0.020110877230763435
step: 160, loss: 0.2698287069797516
step: 170, loss: 0.2003350704908371
step: 180, loss: 0.1285313218832016
step: 190, loss: 0.18118593096733093
step: 200, loss: 0.1540180891752243
step: 210, loss: 0.08826330304145813

```

```

step: 220, loss: 0.11524197459220886
step: 230, loss: 0.1373111605644226
step: 240, loss: 0.0820443257689476
step: 250, loss: 0.0818231999874115
step: 260, loss: 0.01803726889193058
step: 270, loss: 0.022143520414829254
step: 280, loss: 0.10230319201946259
step: 290, loss: 0.12363552302122116
step: 300, loss: 0.1655644029378891
step: 310, loss: 0.058389462530612946
step: 320, loss: 0.21368449926376343
step: 330, loss: 0.1537579745054245
step: 340, loss: 0.2181001454591751
step: 350, loss: 0.13052338361740112
step: 360, loss: 0.06916315108537674
step: 370, loss: 0.1468198597431183
step: 380, loss: 0.1252724677324295
step: 390, loss: 0.04438553377985954
step: 400, loss: 0.011187296360731125
step: 410, loss: 0.033052708953619
step: 420, loss: 0.012571356259286404
step: 430, loss: 0.12250954657793045
step: 440, loss: 0.07528428733348846
step: 450, loss: 0.012716024182736874
step: 460, loss: 0.12062034755945206
step: 470, loss: 0.2395900934934616
step: 480, loss: 0.2805638611316681
step: 490, loss: 0.028164856135845184
step: 500, loss: 0.049527205526828766
step: 510, loss: 0.048179663717746735
step: 520, loss: 0.04215819388628006
step: 530, loss: 0.07203993201255798
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9314179796107508, f1=0.9187935034802784,
best_f1=0.9187935034802784
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07862807810306549
step: 10, loss: 0.2448810189962387
step: 20, loss: 0.10258924216032028
step: 30, loss: 0.060603827238082886

```


step: 40, loss: 0.00805678404867649
step: 50, loss: 0.2379036843776703
step: 60, loss: 0.3788846433162689
step: 70, loss: 0.049654703587293625
step: 80, loss: 0.006861718371510506
step: 90, loss: 0.05018300563097
step: 100, loss: 0.593314528465271
step: 110, loss: 0.030287371948361397
step: 120, loss: 0.03668288514018059
step: 130, loss: 0.0640704408288002
step: 140, loss: 0.024245083332061768
step: 150, loss: 0.05640814080834389
step: 160, loss: 0.00917138997465372
step: 170, loss: 0.04179107025265694
step: 180, loss: 0.007397029548883438
step: 190, loss: 0.07567183673381805
step: 200, loss: 0.03531361743807793
step: 210, loss: 0.013176219537854195
step: 220, loss: 0.006917938590049744
step: 230, loss: 0.01364192459732294
step: 240, loss: 0.1388075202703476
step: 250, loss: 0.0422016903758049
step: 260, loss: 0.045432012528181076
step: 270, loss: 0.10824797302484512
step: 280, loss: 0.07935714721679688
step: 290, loss: 0.08177410066127777
step: 300, loss: 0.00944448821246624
step: 310, loss: 0.09306406229734421
step: 320, loss: 0.04083980619907379
step: 330, loss: 0.013384655117988586
step: 340, loss: 0.01163099892437458
step: 350, loss: 0.046969518065452576
step: 360, loss: 0.040714994072914124
step: 370, loss: 0.02183864265680313
step: 380, loss: 0.056311916559934616
step: 390, loss: 0.014625308103859425
step: 400, loss: 0.05261708050966263
step: 410, loss: 0.0005146716139279306
step: 420, loss: 0.023159144446253777
step: 430, loss: 0.017239151522517204
step: 440, loss: 0.07309966534376144
step: 450, loss: 0.022271933034062386
step: 460, loss: 0.29535698890686035
step: 470, loss: 0.06945591419935226
step: 480, loss: 0.35039597749710083
step: 490, loss: 0.06664693355560303
step: 500, loss: 0.024003952741622925
step: 510, loss: 0.05509845167398453

```

step: 520, loss: 0.040219444781541824
step: 530, loss: 0.121622234582901
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9381443298969072, f1=0.924741298212606,
best_f1=0.924741298212606
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.011405443772673607
step: 10, loss: 0.04809768125414848
step: 20, loss: 0.11840947717428207
step: 30, loss: 0.18580898642539978
step: 40, loss: 0.027573194354772568
step: 50, loss: 0.04423194006085396
step: 60, loss: 0.014351644553244114
step: 70, loss: 0.008266323246061802
step: 80, loss: 0.010371998883783817
step: 90, loss: 0.054544489830732346
step: 100, loss: 0.011854494921863079
step: 110, loss: 0.017012467607855797
step: 120, loss: 0.08471738547086716
step: 130, loss: 0.09060097485780716
step: 140, loss: 0.05462026223540306
step: 150, loss: 0.027407793328166008
step: 160, loss: 0.006851390935480595
step: 170, loss: 0.003227096749469638
step: 180, loss: 0.0027976741548627615
step: 190, loss: 0.10341464728116989
step: 200, loss: 0.035560425370931625
step: 210, loss: 0.11181850731372833
step: 220, loss: 0.06981908529996872
step: 230, loss: 0.02914792113006115
step: 240, loss: 0.027558261528611183
step: 250, loss: 0.007481249049305916
step: 260, loss: 0.009975647553801537
step: 270, loss: 0.009661782532930374
step: 280, loss: 0.02037677727639675
step: 290, loss: 0.015676651149988174
step: 300, loss: 0.15681526064872742
step: 310, loss: 0.12634125351905823
step: 320, loss: 0.12604346871376038
step: 330, loss: 0.016697483137249947

```

```
step: 340, loss: 0.008533566258847713
step: 350, loss: 0.020466310903429985
step: 360, loss: 0.006717306561768055
step: 370, loss: 0.0056869289837777615
step: 380, loss: 0.009720356203615665
step: 390, loss: 0.03169615566730499
step: 400, loss: 0.050417590886354446
step: 410, loss: 0.01732018031179905
step: 420, loss: 0.04362073540687561
step: 430, loss: 0.007475942838937044
step: 440, loss: 0.01578766107559204
step: 450, loss: 0.10039671510457993
step: 460, loss: 0.11676336079835892
step: 470, loss: 0.00677553191781044
step: 480, loss: 0.0021959838923066854
step: 490, loss: 0.11441604793071747
step: 500, loss: 0.15530100464820862
step: 510, loss: 0.005660324357450008
step: 520, loss: 0.01339681539684534
step: 530, loss: 0.027676021680235863
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.9347014925373135, f1=0.9293023255813954,
best_f1=0.924741298212606
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0028910182882100344
step: 10, loss: 0.007041464559733868
step: 20, loss: 0.11506430059671402
step: 30, loss: 0.0030329583678394556
step: 40, loss: 0.005929815117269754
step: 50, loss: 0.018069211393594742
step: 60, loss: 0.002822547685354948
step: 70, loss: 0.00045607343781739473
step: 80, loss: 0.021825727075338364
step: 90, loss: 0.011158685199916363
step: 100, loss: 0.00556902913376689
step: 110, loss: 0.003083752701058984
step: 120, loss: 0.004942185711115599
step: 130, loss: 0.029301609843969345
step: 140, loss: 0.03914879634976387
step: 150, loss: 0.00888601504266262
step: 160, loss: 0.022585242986679077
step: 170, loss: 0.024431519210338593
step: 180, loss: 0.0074647702276706696
step: 190, loss: 0.01637679524719715
```

```

step: 200, loss: 0.002589068841189146
step: 210, loss: 0.04251713678240776
step: 220, loss: 0.0025436850264668465
step: 230, loss: 0.1487709879875183
step: 240, loss: 0.003961410839110613
step: 250, loss: 0.0020358304027467966
step: 260, loss: 0.1940530389547348
step: 270, loss: 0.046213097870349884
step: 280, loss: 0.009801505133509636
step: 290, loss: 0.1033756360411644
step: 300, loss: 0.0014223037287592888
step: 310, loss: 0.002596437931060791
step: 320, loss: 0.14701122045516968
step: 330, loss: 0.08009432256221771
step: 340, loss: 0.10723481327295303
step: 350, loss: 0.05288310348987579
step: 360, loss: 0.13234402239322662
step: 370, loss: 0.04190191254019737
step: 380, loss: 0.00832885317504406
step: 390, loss: 0.03419322893023491
step: 400, loss: 0.011829114519059658
step: 410, loss: 0.05076175183057785
step: 420, loss: 0.011044178158044815
step: 430, loss: 0.027645966038107872
step: 440, loss: 0.11278081685304642
step: 450, loss: 0.03373178467154503
step: 460, loss: 0.023203527554869652
step: 470, loss: 0.009183055721223354
step: 480, loss: 0.0019750637002289295
step: 490, loss: 0.03413589298725128
step: 500, loss: 0.013401337899267673
step: 510, loss: 0.19852173328399658
step: 520, loss: 0.01764153316617012
step: 530, loss: 0.003757557598873973
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 4: dev_f1=0.9366165070679434, f1=0.9223034734917733,
best_f1=0.924741298212606
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002652600407600403
step: 10, loss: 0.010761206969618797
step: 20, loss: 0.0011032032780349255
step: 30, loss: 0.0009360975818708539
step: 40, loss: 0.03342107683420181
step: 50, loss: 0.026595190167427063

```

step: 60, loss: 0.02090306393802166
step: 70, loss: 0.0010196213843300939
step: 80, loss: 0.05566347390413284
step: 90, loss: 0.016030460596084595
step: 100, loss: 0.0008461673860438168
step: 110, loss: 0.0010948569979518652
step: 120, loss: 0.005001155659556389
step: 130, loss: 0.00716310553252697
step: 140, loss: 0.004110421985387802
step: 150, loss: 0.0010397246805951
step: 160, loss: 0.07755757868289948
step: 170, loss: 0.007669944316148758
step: 180, loss: 0.008173931390047073
step: 190, loss: 0.0025353652890771627
step: 200, loss: 0.04513363167643547
step: 210, loss: 0.001361789065413177
step: 220, loss: 0.001350419595837593
step: 230, loss: 0.001531203044578433
step: 240, loss: 0.06575668603181839
step: 250, loss: 0.0015756601933389902
step: 260, loss: 0.007651213090866804
step: 270, loss: 0.010275776498019695
step: 280, loss: 0.18181061744689941
step: 290, loss: 0.004401788115501404
step: 300, loss: 0.02896965853869915
step: 310, loss: 0.023807091638445854
step: 320, loss: 0.008997436612844467
step: 330, loss: 0.010230064392089844
step: 340, loss: 0.020413631573319435
step: 350, loss: 0.025849832221865654
step: 360, loss: 0.09101974964141846
step: 370, loss: 0.06262902170419693
step: 380, loss: 0.038479678332805634
step: 390, loss: 0.05401895195245743
step: 400, loss: 0.0416933074593544
step: 410, loss: 0.00484852772206068
step: 420, loss: 0.00386006198823452
step: 430, loss: 0.10242578387260437
step: 440, loss: 0.023506665602326393
step: 450, loss: 0.0016620971728116274
step: 460, loss: 0.014360679313540459
step: 470, loss: 0.050724249333143234
step: 480, loss: 0.051208607852458954
step: 490, loss: 0.012072033248841763
step: 500, loss: 0.022254589945077896
step: 510, loss: 0.007622361183166504
step: 520, loss: 0.0005887623410671949
step: 530, loss: 0.004781267140060663

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.9382257315373896, f1=0.9253592953175707,
best_f1=0.9253592953175707
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05903277546167374
step: 10, loss: 0.0017961598932743073
step: 20, loss: 0.02550196833908558
step: 30, loss: 0.003694058395922184
step: 40, loss: 0.0027283895760774612
step: 50, loss: 0.0033491449430584908
step: 60, loss: 0.013074621558189392
step: 70, loss: 0.03584093600511551
step: 80, loss: 0.0002929542097263038
step: 90, loss: 0.0018527416978031397
step: 100, loss: 0.13419905304908752
step: 110, loss: 0.05062668398022652
step: 120, loss: 0.025656187906861305
step: 130, loss: 0.0006556300213560462
step: 140, loss: 0.010362952016294003
step: 150, loss: 0.005517014767974615
step: 160, loss: 0.0032424586825072765
step: 170, loss: 0.0010378352599218488
step: 180, loss: 0.0015812524361535907
step: 190, loss: 0.002099971054121852
step: 200, loss: 0.000491700426209718
step: 210, loss: 0.0013318577548488975
step: 220, loss: 0.0014106614980846643
step: 230, loss: 0.005486228037625551
step: 240, loss: 0.0012847783509641886
step: 250, loss: 0.00386811257340014
step: 260, loss: 0.0004559965163934976
step: 270, loss: 0.0041425335220992565
step: 280, loss: 0.015089630149304867
step: 290, loss: 0.0009151000995188951
step: 300, loss: 0.0006600567721761763
step: 310, loss: 0.009806061163544655
step: 320, loss: 0.14112645387649536
step: 330, loss: 0.00228617200627923
step: 340, loss: 0.0009397227549925447
step: 350, loss: 0.1397775262594223

```

```

step: 360, loss: 0.0093567268922925
step: 370, loss: 0.021467437967658043
step: 380, loss: 0.08396956324577332
step: 390, loss: 0.03206389397382736
step: 400, loss: 0.0012788457097485662
step: 410, loss: 0.0008190094958990812
step: 420, loss: 0.019311077892780304
step: 430, loss: 0.0003404678136575967
step: 440, loss: 0.002783149015158415
step: 450, loss: 0.001564282109029591
step: 460, loss: 0.0006103440537117422
step: 470, loss: 0.21696126461029053
step: 480, loss: 0.03571910783648491
step: 490, loss: 0.007340851239860058
step: 500, loss: 0.03185706585645676
step: 510, loss: 0.0012881761649623513
step: 520, loss: 0.022643091157078743
step: 530, loss: 0.011347888968884945
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 6: dev_f1=0.9424964936886395, f1=0.9245194561650257,
best_f1=0.9245194561650257
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.025359567254781723
step: 10, loss: 0.07881411164999008
step: 20, loss: 0.005630070343613625
step: 30, loss: 0.004097561817616224
step: 40, loss: 0.0002291333075845614
step: 50, loss: 0.0045303343795239925
step: 60, loss: 0.0763346254825592
step: 70, loss: 0.06182568520307541
step: 80, loss: 0.012030774727463722
step: 90, loss: 0.0005023808917030692
step: 100, loss: 0.001149983610957861
step: 110, loss: 0.003695845138281584
step: 120, loss: 0.0052485535852611065
step: 130, loss: 0.004319113679230213
step: 140, loss: 0.00311196013353765
step: 150, loss: 0.0003328199090901762
step: 160, loss: 0.0013181124813854694
step: 170, loss: 0.01605233922600746

```

```

step: 180, loss: 0.05676265433430672
step: 190, loss: 0.00021353432384785265
step: 200, loss: 0.0012397660175338387
step: 210, loss: 0.0018482939340174198
step: 220, loss: 0.010419009253382683
step: 230, loss: 0.001143422443419695
step: 240, loss: 0.027283910661935806
step: 250, loss: 0.0050013307482004166
step: 260, loss: 0.0006018480053171515
step: 270, loss: 0.0002820860536303371
step: 280, loss: 0.006885947659611702
step: 290, loss: 0.00043326153536327183
step: 300, loss: 0.0023942512925714254
step: 310, loss: 0.004692560993134975
step: 320, loss: 0.032497305423021317
step: 330, loss: 0.007850231602787971
step: 340, loss: 0.002466689795255661
step: 350, loss: 0.00021830301557201892
step: 360, loss: 0.0050573889166116714
step: 370, loss: 0.0029478345531970263
step: 380, loss: 0.022216537967324257
step: 390, loss: 0.0003957241424359381
step: 400, loss: 0.00010804329212987795
step: 410, loss: 0.002574201673269272
step: 420, loss: 0.0028184722177684307
step: 430, loss: 0.0011148029007017612
step: 440, loss: 0.0003546686202753335
step: 450, loss: 0.027460813522338867
step: 460, loss: 0.0007286721956916153
step: 470, loss: 0.12047938257455826
step: 480, loss: 0.0046875933185219765
step: 490, loss: 0.0023949197493493557
step: 500, loss: 8.739221084397286e-05
step: 510, loss: 0.0015680985525250435
step: 520, loss: 0.0018469917122274637
step: 530, loss: 0.02194487303495407
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9386991109031353, f1=0.9273323956868261,
best_f1=0.9245194561650257
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0012119879247620702
step: 10, loss: 0.011441408656537533
step: 20, loss: 0.02000555768609047
step: 30, loss: 0.0029877140186727047

```


step: 40, loss: 0.00036005806759931147
step: 50, loss: 0.0005283991922624409
step: 60, loss: 0.00582409230992198
step: 70, loss: 0.001892685191705823
step: 80, loss: 0.0015533780679106712
step: 90, loss: 0.0014091001357883215
step: 100, loss: 0.0016302793519571424
step: 110, loss: 0.0004966078558936715
step: 120, loss: 0.003566202474758029
step: 130, loss: 0.0053263320587575436
step: 140, loss: 0.0001348913792753592
step: 150, loss: 0.013530046679079533
step: 160, loss: 0.003738963510841131
step: 170, loss: 0.012404435314238071
step: 180, loss: 0.005121129099279642
step: 190, loss: 0.0032182298600673676
step: 200, loss: 0.009189730510115623
step: 210, loss: 0.00027916007093153894
step: 220, loss: 0.007289613597095013
step: 230, loss: 0.031669095158576965
step: 240, loss: 0.00037174797034822404
step: 250, loss: 0.004594010766595602
step: 260, loss: 0.22317779064178467
step: 270, loss: 0.00010022700735135004
step: 280, loss: 0.001725227339193225
step: 290, loss: 0.02377929911017418
step: 300, loss: 0.0005688757519237697
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
step: 310, loss: 0.029546402394771576
step: 320, loss: 0.0003288400184828788
step: 330, loss: 0.017184654250741005
step: 340, loss: 0.0035460994113236666
step: 350, loss: 0.0022677516099065542
step: 360, loss: 0.009365180507302284
step: 370, loss: 0.002023326465860009
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 380, loss: 0.05140283703804016
step: 390, loss: 9.225239045917988e-05
step: 400, loss: 0.06955866515636444
step: 410, loss: 0.0003346108424011618
step: 420, loss: 0.001795536489225924
step: 430, loss: 0.0021027876064181328
step: 440, loss: 0.005447803530842066
step: 450, loss: 0.0045241923071444035
step: 460, loss: 0.021956024691462517
step: 470, loss: 0.0005699131870642304
step: 480, loss: 0.0005401863600127399
step: 490, loss: 0.0006821950664743781

```

step: 500, loss: 0.0003576503659132868
step: 510, loss: 0.0004181156400591135
step: 520, loss: 0.00031000745366327465
step: 530, loss: 6.639458297286183e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9364426154549611, f1=0.9322419281491587,
best_f1=0.9245194561650257
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0020340741612017155
step: 10, loss: 0.007494909223169088
step: 20, loss: 0.00012054628314217553
step: 30, loss: 0.0004966150736436248
step: 40, loss: 0.0021119751036167145
step: 50, loss: 0.0017672603717073798
step: 60, loss: 0.00020367438264656812
step: 70, loss: 0.1752818524837494
step: 80, loss: 0.002942396327853203
step: 90, loss: 0.010592171922326088
step: 100, loss: 0.0004013717407360673
step: 110, loss: 0.0006926728528924286
step: 120, loss: 7.910963904578239e-05
step: 130, loss: 0.00039351731538772583
step: 140, loss: 0.012647133320569992
step: 150, loss: 0.0010621044784784317
step: 160, loss: 0.020491130650043488
step: 170, loss: 0.0009380051051266491
step: 180, loss: 0.002207694575190544
step: 190, loss: 0.0016935719177126884
step: 200, loss: 0.0005742807406932116
step: 210, loss: 0.00015876724501140416
step: 220, loss: 0.009933291003108025
step: 230, loss: 0.0003075697459280491
step: 240, loss: 0.06364724785089493
step: 250, loss: 0.0064890035428106785
step: 260, loss: 0.0009068602812476456
step: 270, loss: 0.03908490017056465
step: 280, loss: 0.0005811613518744707
step: 290, loss: 7.194092177087441e-05
step: 300, loss: 0.020537985488772392
step: 310, loss: 0.0003327852173242718
step: 320, loss: 0.0008492773631587625
step: 330, loss: 0.0049575199373066425
step: 340, loss: 0.00041428758413530886
step: 350, loss: 4.0967748645925894e-05

```

```
step: 360, loss: 0.024253977462649345
step: 370, loss: 0.0002703357895370573
step: 380, loss: 0.0004106323467567563
step: 390, loss: 0.0018548330990597606
step: 400, loss: 0.003562341211363673
step: 410, loss: 0.0017980252159759402
step: 420, loss: 0.0036696726456284523
step: 430, loss: 0.0008337526232935488
step: 440, loss: 0.0002447425213176757
step: 450, loss: 0.00011380253272363916
step: 460, loss: 0.0043387641198933125
step: 470, loss: 7.919955532997847e-05
step: 480, loss: 0.03995329886674881
step: 490, loss: 0.0014021407114341855
step: 500, loss: 0.0046415249817073345
step: 510, loss: 0.004520382732152939
step: 520, loss: 0.0007644033757969737
step: 530, loss: 0.0005946432356722653
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.9397031539888682, f1=0.9297597042513862,
best_f1=0.9245194561650257
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00047019473277032375
step: 10, loss: 0.0006575066363438964
step: 20, loss: 0.0007514860481023788
step: 30, loss: 0.00026342025375925004
step: 40, loss: 0.0003390319470781833
step: 50, loss: 0.00041585348662920296
step: 60, loss: 0.0005541478167288005
step: 70, loss: 0.0003390591882634908
step: 80, loss: 0.0018300339579582214
step: 90, loss: 0.0009218686609528959
step: 100, loss: 0.00021022239525336772
step: 110, loss: 0.005698116961866617
step: 120, loss: 0.0038841371424496174
step: 130, loss: 0.0008268934325315058
step: 140, loss: 0.011106341145932674
step: 150, loss: 0.001366417040117085
step: 160, loss: 0.00020122983551118523
step: 170, loss: 0.00012443639570847154
step: 180, loss: 0.0006241974770091474
step: 190, loss: 0.0006289836601354182
step: 200, loss: 0.0006301229586824775
step: 210, loss: 0.011794522404670715
```

```

step: 220, loss: 0.00017726562509778887
step: 230, loss: 0.006539300084114075
step: 240, loss: 0.00011225610796827823
step: 250, loss: 0.0017278279410675168
step: 260, loss: 0.005016733426600695
step: 270, loss: 0.014560014940798283
step: 280, loss: 8.366215479327366e-05
step: 290, loss: 0.0002139141724910587
step: 300, loss: 0.003633287502452731
step: 310, loss: 0.0014783097431063652
step: 320, loss: 0.000762492825742811
step: 330, loss: 0.03151237592101097
step: 340, loss: 0.0036281822249293327
step: 350, loss: 0.0003603047225624323
step: 360, loss: 0.11183150112628937
step: 370, loss: 0.0018969104858115315
step: 380, loss: 0.0007321890443563461
step: 390, loss: 0.0032283980399370193
step: 400, loss: 0.00022572641319129616
step: 410, loss: 0.0013352378737181425
step: 420, loss: 0.007122100330889225
step: 430, loss: 0.0005556316464208066
step: 440, loss: 5.509893162525259e-05
step: 450, loss: 0.002034986624494195
step: 460, loss: 0.003903480712324381
step: 470, loss: 0.004468420520424843
step: 480, loss: 0.00012437812983989716
step: 490, loss: 0.11309937387704849
step: 500, loss: 0.009106173180043697
step: 510, loss: 0.0012722766259685159
step: 520, loss: 0.001976157771423459
step: 530, loss: 0.00026576293748803437
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9408476944573824, f1=0.9263059701492538,
best_f1=0.9245194561650257
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007204187568277121
step: 10, loss: 0.020334484055638313
step: 20, loss: 0.00014737321180291474
step: 30, loss: 0.002168460050597787
step: 40, loss: 0.0017268622759729624
step: 50, loss: 0.000868881237693131
step: 60, loss: 0.00012561996118165553
step: 70, loss: 0.00019261642592027783

```

step: 80, loss: 0.0009174797451123595
step: 90, loss: 0.00016437339945696294
step: 100, loss: 0.04791172221302986
step: 110, loss: 6.858304550405592e-05
step: 120, loss: 0.0035944648552685976
step: 130, loss: 0.00039861424011178315
step: 140, loss: 0.00034529916592873633
step: 150, loss: 8.89761358848773e-05
step: 160, loss: 0.000300604006042704
step: 170, loss: 0.0014194308314472437
step: 180, loss: 0.002657664241269231
step: 190, loss: 0.007170215714722872
step: 200, loss: 0.0017520005349069834
step: 210, loss: 0.0003226306871511042
step: 220, loss: 0.0001797234290279448
step: 230, loss: 0.002550754463300109
step: 240, loss: 3.734299025381915e-05
step: 250, loss: 0.0040734559297561646
step: 260, loss: 6.570298137376085e-05
step: 270, loss: 0.00911071989685297
step: 280, loss: 0.0006415965035557747
step: 290, loss: 0.003728361101821065
step: 300, loss: 0.07488391548395157
step: 310, loss: 0.000494243751745671
step: 320, loss: 0.016454650089144707
step: 330, loss: 0.000274160411208868
step: 340, loss: 0.022284282371401787
step: 350, loss: 0.01094890758395195
step: 360, loss: 0.0009303808910772204
step: 370, loss: 4.399566023494117e-05
step: 380, loss: 3.107756492681801e-05
step: 390, loss: 0.002773730084300041
step: 400, loss: 2.2098007320892066e-05
step: 410, loss: 0.00029373011784628034
step: 420, loss: 0.009783502668142319
step: 430, loss: 0.00014657182327937335
step: 440, loss: 0.0003462902386672795
step: 450, loss: 0.00022708009055349976
step: 460, loss: 0.0025564224924892187
step: 470, loss: 0.0005704189534299076
step: 480, loss: 0.0004656244127545506
step: 490, loss: 0.00024250328715424985
step: 500, loss: 0.00011305360385449603
step: 510, loss: 3.118675886071287e-05
step: 520, loss: 3.0117575079202652e-05
step: 530, loss: 4.815891952603124e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9387379087977891, f1=0.926829268292683,
best_f1=0.9245194561650257
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00018784466374199837
step: 10, loss: 0.0003344374999869615
step: 20, loss: 2.0257879441487603e-05
step: 30, loss: 6.260148802539334e-05
step: 40, loss: 0.00011701785842888057
step: 50, loss: 0.0006945943459868431
step: 60, loss: 0.00013657976523973048
step: 70, loss: 0.038409747183322906
step: 80, loss: 0.0005690547986887395
step: 90, loss: 2.930977825599257e-05
step: 100, loss: 0.00043274147901684046
step: 110, loss: 0.06388113647699356
step: 120, loss: 0.0031713638454675674
step: 130, loss: 0.0001315269764745608
step: 140, loss: 8.755358430789784e-05
step: 150, loss: 0.013698062859475613
step: 160, loss: 0.00033985363552346826
step: 170, loss: 0.0007167900912463665
step: 180, loss: 0.000281330052530393
step: 190, loss: 0.00021552681573666632
step: 200, loss: 7.819895108696073e-05
step: 210, loss: 0.00018350612663198262
step: 220, loss: 0.007205740548670292
step: 230, loss: 0.006158497184514999
step: 240, loss: 4.7753761464264244e-05
step: 250, loss: 0.00011174652172485366
step: 260, loss: 0.0005385131807997823
step: 270, loss: 0.0003610718995332718
step: 280, loss: 5.728731775889173e-05
step: 290, loss: 0.0036430037580430508
step: 300, loss: 0.34362977743148804
step: 310, loss: 0.0002431506582070142
step: 320, loss: 4.6495228161802515e-05
step: 330, loss: 0.009270875714719296
step: 340, loss: 0.0030543070752173662
step: 350, loss: 0.02461261674761772
step: 360, loss: 0.0009161181515082717
step: 370, loss: 0.006739444099366665
step: 380, loss: 0.001782681210897863
step: 390, loss: 0.00014682119945064187
step: 400, loss: 0.0003949040255974978
step: 410, loss: 0.009830745868384838

```

```

step: 420, loss: 0.0020674290135502815
step: 430, loss: 0.0003237615746911615
step: 440, loss: 0.0007793292170390487
step: 450, loss: 0.0005293659050948918
step: 460, loss: 0.00013835320714861155
step: 470, loss: 0.0011361913057044148
step: 480, loss: 0.0012166526867076755
step: 490, loss: 0.007718062959611416
step: 500, loss: 0.00453055277466774
step: 510, loss: 0.0011454280465841293
step: 520, loss: 0.017990238964557648
step: 530, loss: 0.0003453076642472297
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9382022471910113, f1=0.9260991580916744,
best_f1=0.9245194561650257
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004917437909170985
step: 10, loss: 0.00048414297634735703
step: 20, loss: 0.026796726509928703
step: 30, loss: 0.00027730027795769274
step: 40, loss: 0.010674888268113136
step: 50, loss: 0.00014319171896204352
step: 60, loss: 0.0007065105019137263
step: 70, loss: 5.579981370829046e-05
step: 80, loss: 0.002993933390825987
step: 90, loss: 0.004839069209992886
step: 100, loss: 1.920361137308646e-05
step: 110, loss: 0.005801440216600895
step: 120, loss: 0.0023043924011290073
step: 130, loss: 0.0012561125913634896
step: 140, loss: 6.578659667866305e-05
step: 150, loss: 9.876703552436084e-05
step: 160, loss: 0.001582612283527851
step: 170, loss: 0.03784705698490143
step: 180, loss: 0.00017234127153642476
step: 190, loss: 0.00023629739007446915
step: 200, loss: 0.006586235016584396
step: 210, loss: 0.0003633700544014573
step: 220, loss: 0.0037800956051796675
step: 230, loss: 0.0001878183102235198
step: 240, loss: 4.899632040178403e-05
step: 250, loss: 0.037132300436496735
step: 260, loss: 0.000690927729010582
step: 270, loss: 0.004658011253923178

```

```
step: 280, loss: 0.0004704984021373093
step: 290, loss: 8.571194484829903e-05
step: 300, loss: 0.0335482694208622
step: 310, loss: 0.0011598862474784255
step: 320, loss: 4.02605946874246e-05
step: 330, loss: 0.0012570064282044768
step: 340, loss: 4.851467383559793e-05
step: 350, loss: 0.04099955037236214
step: 360, loss: 0.00035034719621762633
step: 370, loss: 0.038036245852708817
step: 380, loss: 0.0006544435163959861
step: 390, loss: 0.0005967033794149756
step: 400, loss: 6.823386502219364e-05
step: 410, loss: 0.000298069731798023
step: 420, loss: 0.00043053386616520584
step: 430, loss: 0.0014058081433176994
step: 440, loss: 0.019108938053250313
step: 450, loss: 0.000202534458367154
step: 460, loss: 0.0008440700476057827
step: 470, loss: 0.017118126153945923
step: 480, loss: 0.0010316979605704546
step: 490, loss: 0.004771031439304352
step: 500, loss: 0.0003891604719683528
step: 510, loss: 4.242609429638833e-05
step: 520, loss: 0.0035871996078640223
step: 530, loss: 0.00019239769608248025
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.9370300751879698, f1=0.9341429238673518,
best_f1=0.9245194561650257
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00028924233629368246
step: 10, loss: 0.013844079338014126
step: 20, loss: 4.9892412789631635e-05
step: 30, loss: 0.0008640136220492423
step: 40, loss: 0.0014389909338206053
step: 50, loss: 0.0016746658366173506
step: 60, loss: 7.782434113323689e-05
step: 70, loss: 6.991544796619564e-05
step: 80, loss: 0.002013257471844554
step: 90, loss: 0.00014003309479448944
step: 100, loss: 0.02224113419651985
step: 110, loss: 4.790123784914613e-05
step: 120, loss: 0.0001006598467938602
step: 130, loss: 0.005612103268504143
```



```

step: 140, loss: 0.0022656102664768696
step: 150, loss: 0.000489931320771575
step: 160, loss: 0.00026208258350379765
step: 170, loss: 0.00013555781333707273
step: 180, loss: 0.0013641436817124486
step: 190, loss: 0.0005524471052922308
step: 200, loss: 0.0007246396271511912
step: 210, loss: 0.0013618910452350974
step: 220, loss: 0.05885741114616394
step: 230, loss: 0.00014042116526979953
step: 240, loss: 0.00034274670179001987
step: 250, loss: 2.99050170724513e-05
step: 260, loss: 2.8514827135950327e-05
step: 270, loss: 0.0018227703403681517
step: 280, loss: 0.0008339406922459602
step: 290, loss: 8.970589260570705e-05
step: 300, loss: 0.019155053421854973
step: 310, loss: 0.0020709671080112457
step: 320, loss: 0.00046017864951863885
step: 330, loss: 0.0006704629631713033
step: 340, loss: 6.401228165486827e-05
step: 350, loss: 0.0003212711017113179
step: 360, loss: 0.0015044074971228838
step: 370, loss: 0.0012925644405186176
step: 380, loss: 0.0011493603233247995
step: 390, loss: 0.00020606277394108474
step: 400, loss: 3.451671727816574e-05
step: 410, loss: 8.643321780255064e-05
step: 420, loss: 0.0003165118978358805
step: 430, loss: 7.755253318464383e-05
step: 440, loss: 3.359649781486951e-05
step: 450, loss: 0.0011415887856855989
step: 460, loss: 0.0197439081966877
step: 470, loss: 0.0001826303341658786
step: 480, loss: 3.180554631398991e-05
step: 490, loss: 0.0001316508714808151
step: 500, loss: 6.051461605238728e-05
step: 510, loss: 0.0001478650519857183
step: 520, loss: 3.4053016861435026e-05
step: 530, loss: 0.00025281342095695436
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.938568129330254, f1=0.9362880886426593,
best_f1=0.9245194561650257
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```

step: 0, loss: 0.00011735917360056192
step: 10, loss: 0.0005747458199039102
step: 20, loss: 0.0003334824286866933
step: 30, loss: 0.0029199779964983463
step: 40, loss: 5.224963024375029e-05
step: 50, loss: 0.0009293664479628205
step: 60, loss: 0.00010135625780094415
step: 70, loss: 0.00011275997530901805
step: 80, loss: 0.0008128179470077157
step: 90, loss: 1.3958555427961983e-05
step: 100, loss: 4.710086795967072e-05
step: 110, loss: 0.015055738389492035
step: 120, loss: 0.000519328226801008
step: 130, loss: 0.0001200288679683581
step: 140, loss: 0.0001709337520878762
step: 150, loss: 0.004825425334274769
step: 160, loss: 0.0001625627774046734
step: 170, loss: 0.00040750199696049094
step: 180, loss: 0.0021409434266388416
step: 190, loss: 0.017712567001581192
step: 200, loss: 0.002630312228575349
step: 210, loss: 0.0012429316993802786
step: 220, loss: 0.00038665730971843004
step: 230, loss: 0.00037039187736809254
step: 240, loss: 7.433331484207883e-05
step: 250, loss: 0.0016242660349234939
step: 260, loss: 0.0009837574325501919
step: 270, loss: 0.0101197874173522
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 280, loss: 0.001330118509940803
step: 290, loss: 0.0003885850019287318
step: 300, loss: 0.004377901088446379
step: 310, loss: 0.0009490789962001145
step: 320, loss: 0.00010174392082262784
step: 330, loss: 0.0039171441458165646
step: 340, loss: 0.00037076385342516005
step: 350, loss: 0.006520211696624756
step: 360, loss: 2.6832352887140587e-05
step: 370, loss: 4.646586603485048e-05
step: 380, loss: 5.5309188610408455e-05
step: 390, loss: 0.00010256397217744961
step: 400, loss: 0.0014905721182003617
step: 410, loss: 0.0008940154802985489
step: 420, loss: 0.006602742709219456
step: 430, loss: 9.223398956237361e-05
step: 440, loss: 0.003876610193401575
step: 450, loss: 0.004343628883361816
step: 460, loss: 4.9050708184950054e-05

```

step: 470, loss: 0.00014765663945581764
step: 480, loss: 0.00016366061754524708
step: 490, loss: 0.00021976357675157487
step: 500, loss: 0.000656646559946239
step: 510, loss: 1.8987542716786265e-05
step: 520, loss: 0.0002554669335950166
step: 530, loss: 0.0006596117164008319
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9366494603472549, f1=0.9341429238673518,
best_f1=0.9245194561650257
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
5742it [00:22, 256.02it/s]
load_f1 = 0.9385265133740028
real_f1 = 0.9390243902439025
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

4398it [00:18, 242.15it/s]

4.1.4 Amazon-Google - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Structured/Amazon-Google \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model \
  --dk general \
  --dk product \
  --summarize \
  --da entry_swap \
  --da attr_del

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Amazon-Google.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/Amazon-Google \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

enabled	: True
opt_level	: 02
cast_model_type	: torch.float16
patch_torch_functions	: False

```

keep_batchnorm_fp32      : True
master_weights           : True
loss_scale                : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                   : True
opt_level                 : 02
cast_model_type           : torch.float16
patch_torch_functions     : False
keep_batchnorm_fp32      : True
master_weights           : True
loss_scale                : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8365583419799805
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.06360248476266861
step: 20, loss: 0.38095831871032715
step: 30, loss: 0.374766081571579
step: 40, loss: 0.5057409405708313
step: 50, loss: 0.2992970645427704
step: 60, loss: 0.35735347867012024
step: 70, loss: 0.21627767384052277
step: 80, loss: 0.2940734922885895
step: 90, loss: 0.4100753664970398
step: 100, loss: 0.165373295545578
step: 110, loss: 0.2884036898612976
step: 120, loss: 0.22758544981479645
step: 130, loss: 0.1942763328552246
step: 140, loss: 0.15581655502319336
step: 150, loss: 0.4265078902244568
step: 160, loss: 0.2030981481075287
step: 170, loss: 0.11732804030179977
step: 180, loss: 0.1836661398410797

```

```

step: 190, loss: 0.15390510857105255
step: 200, loss: 0.0978865996003151
step: 210, loss: 0.4192832112312317
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.6100519930675911, f1=0.6088495575221239,
best_f1=0.6088495575221239
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09051131457090378
step: 10, loss: 0.03436034545302391
step: 20, loss: 0.2561120092868805
step: 30, loss: 0.10901427268981934
step: 40, loss: 0.08221160620450974
step: 50, loss: 0.20866452157497406
step: 60, loss: 0.06224021688103676
step: 70, loss: 0.14206930994987488
step: 80, loss: 0.25079068541526794
step: 90, loss: 0.19207929074764252
step: 100, loss: 0.0849645733833313
step: 110, loss: 0.12033435702323914
step: 120, loss: 0.12896712124347687
step: 130, loss: 0.1964188516139984
step: 140, loss: 0.21379250288009644
step: 150, loss: 0.2139924317598343
step: 160, loss: 0.10588903725147247
step: 170, loss: 0.20403148233890533
step: 180, loss: 0.17208822071552277
step: 190, loss: 0.12822546064853668
step: 200, loss: 0.329245001077652
step: 210, loss: 0.18901468813419342
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.6486486486486486, f1=0.6703096539162113,
best_f1=0.6703096539162113
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.12311532348394394
step: 10, loss: 0.22216053307056427
step: 20, loss: 0.27695733308792114
step: 30, loss: 0.17686375975608826
step: 40, loss: 0.278364360332489
step: 50, loss: 0.10416955500841141
step: 60, loss: 0.2968166470527649
step: 70, loss: 0.13048818707466125
step: 80, loss: 0.026254260912537575
step: 90, loss: 0.10378718376159668
step: 100, loss: 0.055478788912296295
step: 110, loss: 0.12410373985767365
step: 120, loss: 0.23406453430652618
step: 130, loss: 0.1102638840675354
step: 140, loss: 0.2022978961467743
step: 150, loss: 0.2318820059299469
step: 160, loss: 0.24879072606563568
step: 170, loss: 0.22226952016353607
step: 180, loss: 0.06138324365019798
step: 190, loss: 0.2292930632829666
step: 200, loss: 0.33290091156959534
step: 210, loss: 0.17752023041248322
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.6726943942133816, f1=0.6871609403254972,
best_f1=0.6871609403254972
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.18094541132450104
step: 10, loss: 0.03802113980054855
step: 20, loss: 0.1691480576992035
step: 30, loss: 0.09528093039989471
step: 40, loss: 0.07120814174413681
step: 50, loss: 0.0974632203578949
step: 60, loss: 0.06395930796861649
step: 70, loss: 0.24042664468288422
step: 80, loss: 0.1629372090101242
step: 90, loss: 0.07876703888177872
step: 100, loss: 0.1279694139957428
step: 110, loss: 0.11129479855298996
step: 120, loss: 0.08671609312295914

```

```

step: 130, loss: 0.10607726126909256
step: 140, loss: 0.1499868929386139
step: 150, loss: 0.17853686213493347
step: 160, loss: 0.03884465619921684
step: 170, loss: 0.0713612362742424
step: 180, loss: 0.03608498349785805
step: 190, loss: 0.0497051365673542
step: 200, loss: 0.04799828678369522
step: 210, loss: 0.03576255217194557
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.6790123456790124, f1=0.6624737945492662,
best_f1=0.6624737945492662
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.036246396601200104
step: 10, loss: 0.12139453738927841
step: 20, loss: 0.04172685742378235
step: 30, loss: 0.05700298771262169
step: 40, loss: 0.20812943577766418
step: 50, loss: 0.12377562373876572
step: 60, loss: 0.04149941727519035
step: 70, loss: 0.04708059877157211
step: 80, loss: 0.007454120088368654
step: 90, loss: 0.18063049018383026
step: 100, loss: 0.07722494751214981
step: 110, loss: 0.006461014039814472
step: 120, loss: 0.1273573338985443
step: 130, loss: 0.04214150086045265
step: 140, loss: 0.1071828305721283
step: 150, loss: 0.06575440615415573
step: 160, loss: 0.11397621035575867
step: 170, loss: 0.2150115817785263
step: 180, loss: 0.2290252447128296
step: 190, loss: 0.06759129464626312
step: 200, loss: 0.0698554739356041
step: 210, loss: 0.048488132655620575
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or

```



```

loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.689655172413793, f1=0.7063492063492063,
best_f1=0.7063492063492063
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06048817187547684
step: 10, loss: 0.0888606533408165
step: 20, loss: 0.045734748244285583
step: 30, loss: 0.18538525700569153
step: 40, loss: 0.012631445191800594
step: 50, loss: 0.11627823859453201
step: 60, loss: 0.05577646940946579
step: 70, loss: 0.0035117212682962418
step: 80, loss: 0.08650635182857513
step: 90, loss: 0.07520978897809982
step: 100, loss: 0.03126481547951698
step: 110, loss: 0.07650620490312576
step: 120, loss: 0.017085116356611252
step: 130, loss: 0.05540991201996803
step: 140, loss: 0.09215103834867477
step: 150, loss: 0.09211172163486481
step: 160, loss: 0.10963628441095352
step: 170, loss: 0.06268084794282913
step: 180, loss: 0.018248731270432472
step: 190, loss: 0.14043207466602325
step: 200, loss: 0.007340786978602409
step: 210, loss: 0.023758865892887115
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.6859344894026974, f1=0.7000000000000001,
best_f1=0.7063492063492063
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.020493080839514732
step: 10, loss: 0.021346179768443108
step: 20, loss: 0.024480178952217102
step: 30, loss: 0.012281346134841442
step: 40, loss: 0.053548041731119156
step: 50, loss: 0.04206036403775215
step: 60, loss: 0.1061689704656601
step: 70, loss: 0.005625004414469004
step: 80, loss: 0.07862269133329391
step: 90, loss: 0.013975819572806358
step: 100, loss: 0.016712302342057228

```

```

step: 110, loss: 0.029216086491942406
step: 120, loss: 0.07473713904619217
step: 130, loss: 0.004728004336357117
step: 140, loss: 0.0031682762783020735
step: 150, loss: 0.09296466410160065
step: 160, loss: 0.03101714700460434
step: 170, loss: 0.04798039421439171
step: 180, loss: 0.007775777019560337
step: 190, loss: 0.06908247619867325
step: 200, loss: 0.06825070083141327
step: 210, loss: 0.0715632438659668
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.672, f1=0.6812749003984064, best_f1=0.7063492063492063
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06601358205080032
step: 10, loss: 0.04283100739121437
step: 20, loss: 0.19261467456817627
step: 30, loss: 0.030286533758044243
step: 40, loss: 0.056389570236206055
step: 50, loss: 0.014321526512503624
step: 60, loss: 0.20183812081813812
step: 70, loss: 0.009835230186581612
step: 80, loss: 0.008957869373261929
step: 90, loss: 0.12222645431756973
step: 100, loss: 0.044789500534534454
step: 110, loss: 0.010487343184649944
step: 120, loss: 0.08266275376081467
step: 130, loss: 0.010433412156999111
step: 140, loss: 0.007036691065877676
step: 150, loss: 0.0938073918223381
step: 160, loss: 0.12965761125087738
step: 170, loss: 0.058904729783535004
step: 180, loss: 0.12532413005828857
step: 190, loss: 0.05083765834569931
step: 200, loss: 0.069891057908535
step: 210, loss: 0.129003643989563
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.6666666666666666, f1=0.6548323471400394,
best_f1=0.7063492063492063
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.002343808300793171
step: 10, loss: 0.03202464058995247
step: 20, loss: 0.05640657618641853
step: 30, loss: 0.21235701441764832
step: 40, loss: 0.021725425496697426
step: 50, loss: 0.0788072869181633
step: 60, loss: 0.09061646461486816
step: 70, loss: 0.16836713254451752
step: 80, loss: 0.003876596922054887
step: 90, loss: 0.07449867576360703
step: 100, loss: 0.03040524385869503
step: 110, loss: 0.050598517060279846
step: 120, loss: 0.008917830884456635
step: 130, loss: 0.002455266658216715
step: 140, loss: 0.033187732100486755
step: 150, loss: 0.02851800248026848
step: 160, loss: 0.004708214662969112
step: 170, loss: 0.013950672931969166
step: 180, loss: 0.007732616271823645
step: 190, loss: 0.12060949206352234
step: 200, loss: 0.015486921183764935
step: 210, loss: 0.10760742425918579
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.6794055201698513, f1=0.6822033898305084,
best_f1=0.7063492063492063
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010422132909297943
step: 10, loss: 0.00444313557818532
step: 20, loss: 0.035762351006269455
step: 30, loss: 0.13253745436668396
step: 40, loss: 0.12974339723587036
step: 50, loss: 0.007361721713095903
step: 60, loss: 0.08712749183177948
step: 70, loss: 0.0693911463022232
step: 80, loss: 0.001712522585876286
step: 90, loss: 0.09173917025327682
step: 100, loss: 0.002834015991538763
step: 110, loss: 0.005336098838597536
step: 120, loss: 0.03119025193154812
step: 130, loss: 0.08723028749227524
step: 140, loss: 0.0014632774982601404
step: 150, loss: 0.008408286608755589
step: 160, loss: 0.06205126643180847
step: 170, loss: 0.025053726509213448

```

```

step: 180, loss: 0.014663438312709332
step: 190, loss: 0.28039854764938354
step: 200, loss: 0.08753852546215057
step: 210, loss: 0.07176229357719421
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.6601941747572815, f1=0.6615087040618955,
best_f1=0.7063492063492063
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.013076805509626865
step: 10, loss: 0.02232201397418976
step: 20, loss: 0.07678015530109406
step: 30, loss: 0.0422699972987175
step: 40, loss: 0.03442727029323578
step: 50, loss: 0.03177909180521965
step: 60, loss: 0.008582442998886108
step: 70, loss: 0.0028182859532535076
step: 80, loss: 0.0629502609372139
step: 90, loss: 0.13954642415046692
step: 100, loss: 0.0047111790627241135
step: 110, loss: 0.06821294128894806
step: 120, loss: 0.03836462274193764
step: 130, loss: 0.031250130385160446
step: 140, loss: 0.03137805685400963
step: 150, loss: 0.10609614849090576
step: 160, loss: 0.00302134919911623
step: 170, loss: 0.06665514409542084
step: 180, loss: 0.05268895998597145
step: 190, loss: 0.04224603995680809
step: 200, loss: 0.0005517237004823983
step: 210, loss: 0.005530774127691984
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.6818181818181819, f1=0.672, best_f1=0.7063492063492063
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006399391684681177
step: 10, loss: 0.003934372216463089
step: 20, loss: 0.03115617297589779
step: 30, loss: 0.0072706411592662334
step: 40, loss: 0.04079766571521759
step: 50, loss: 0.06418584287166595
step: 60, loss: 0.022472411394119263

```

```

step: 70, loss: 0.0007011842099018395
step: 80, loss: 0.020497966557741165
step: 90, loss: 0.007537417113780975
step: 100, loss: 0.0004655868688132614
step: 110, loss: 0.004136147443205118
step: 120, loss: 0.0012443872401490808
step: 130, loss: 0.03736502304673195
step: 140, loss: 0.07257165014743805
step: 150, loss: 0.004487596917897463
step: 160, loss: 0.014550822786986828
step: 170, loss: 0.007365578785538673
step: 180, loss: 0.012974599376320839
step: 190, loss: 0.18100163340568542
step: 200, loss: 0.07365966588258743
step: 210, loss: 0.014079231768846512
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.6794871794871795, f1=0.673866090712743,
best_f1=0.7063492063492063
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005050254054367542
step: 10, loss: 0.01081059966236353
step: 20, loss: 0.0026765880174934864
step: 30, loss: 0.020298223942518234
step: 40, loss: 0.005486771930009127
step: 50, loss: 0.01165221817791462
step: 60, loss: 0.0011217250721529126
step: 70, loss: 0.045813314616680145
step: 80, loss: 0.017293009907007217
step: 90, loss: 0.00973600521683693
step: 100, loss: 0.041155677288770676
step: 110, loss: 0.0014294497668743134
step: 120, loss: 0.03501178324222565
step: 130, loss: 0.0015810730401426554
step: 140, loss: 0.02848583459854126
step: 150, loss: 0.0006555979489348829
step: 160, loss: 0.003004831727594137
step: 170, loss: 0.007042011711746454
step: 180, loss: 0.06031707674264908
step: 190, loss: 0.0018876563990488648
step: 200, loss: 0.049158062785863876
step: 210, loss: 0.00563783198595047
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

epoch 13: dev_f1=0.673960612691466, f1=0.6809421841541756,
best_f1=0.7063492063492063
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009053683024831116
step: 10, loss: 0.003773102769628167
step: 20, loss: 0.002560561755672097
step: 30, loss: 0.07270701974630356
step: 40, loss: 0.06516893953084946
step: 50, loss: 0.015655795112252235
step: 60, loss: 0.014846480451524258
step: 70, loss: 0.05753650888800621
step: 80, loss: 0.03193230554461479
step: 90, loss: 0.12307158857584
step: 100, loss: 0.013416489586234093
step: 110, loss: 0.009978658519685268
step: 120, loss: 0.002654947806149721
step: 130, loss: 0.13706764578819275
step: 140, loss: 0.0007158382213674486
step: 150, loss: 0.010227403603494167
step: 160, loss: 0.027008699253201485
step: 170, loss: 0.06970962882041931
step: 180, loss: 0.007633560802787542
step: 190, loss: 0.011361801065504551
step: 200, loss: 0.002323511056602001
step: 210, loss: 0.06790953874588013
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.6747967479674798, f1=0.6814516129032259,
best_f1=0.7063492063492063
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008744711056351662
step: 10, loss: 0.014563548378646374
step: 20, loss: 0.002044541761279106
step: 30, loss: 0.002229852369055152
step: 40, loss: 0.0026427784468978643
step: 50, loss: 0.002089365851134062
step: 60, loss: 0.029167940840125084
step: 70, loss: 0.001454790006391704
step: 80, loss: 0.001101419678889215
step: 90, loss: 0.005931568797677755
step: 100, loss: 0.002036304445937276
step: 110, loss: 0.00048256805166602135
step: 120, loss: 0.02014949731528759

```

```

step: 130, loss: 0.14107859134674072
step: 140, loss: 0.010506844148039818
step: 150, loss: 0.0032533258199691772
step: 160, loss: 0.0035131247714161873
step: 170, loss: 0.022803712636232376
step: 180, loss: 0.0055452329106628895
step: 190, loss: 0.08476383984088898
step: 200, loss: 0.004344947170466185
step: 210, loss: 0.03319530189037323
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.6694736842105263, f1=0.6694560669456067,
best_f1=0.7063492063492063
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
2293it [00:06, 344.08it/s]
load_f1 = 0.6800804828973844
real_f1 = 0.6828282828282828
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-

```

```
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:18, 241.04it/s]
```

4.1.5 Walmart-Amazon - Running the matcher

```
[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Structured/Walmart-Amazon \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model \
  --dk general \
  --dk product \
  --summarize \
  --da entry_swap \
  --da attr_del

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Walmart-Amazon.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Structured/Walmart-Amazon \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/
```

Selected optimization level O2: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```
enabled          : True
opt_level        : O2
```



```

cast_model_type      : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32  : True
master_weights       : True
loss_scale            : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled              : True
opt_level            : 02
cast_model_type      : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32  : True
master_weights       : True
loss_scale            : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8535335659980774
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.1656191647052765
step: 20, loss: 0.1525445282459259
step: 30, loss: 0.5060791373252869
step: 40, loss: 0.25797733664512634
step: 50, loss: 0.3092449903488159
step: 60, loss: 0.3638589084148407
step: 70, loss: 0.16548408567905426
step: 80, loss: 0.5288596749305725
step: 90, loss: 0.23990243673324585
step: 100, loss: 0.21933478116989136
step: 110, loss: 0.23610107600688934
step: 120, loss: 0.40658289194107056
step: 130, loss: 0.3591557741165161
step: 140, loss: 0.27731621265411377
step: 150, loss: 0.2618725001811981
step: 160, loss: 0.19186237454414368

```

```

step: 170, loss: 0.30609002709388733
step: 180, loss: 0.23859147727489471
step: 190, loss: 0.09166442602872849
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.6761363636363635, f1=0.6647058823529411,
best_f1=0.6647058823529411
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.26378777623176575
step: 10, loss: 0.017967993393540382
step: 20, loss: 0.1106710284948349
step: 30, loss: 0.09879869967699051
step: 40, loss: 0.44353729486465454
step: 50, loss: 0.2663723826408386
step: 60, loss: 0.11568750441074371
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 70, loss: 0.17968541383743286
step: 80, loss: 0.1954517364501953
step: 90, loss: 0.0880143865942955
step: 100, loss: 0.2805841267108917
step: 110, loss: 0.11558598279953003
step: 120, loss: 0.08708348125219345
step: 130, loss: 0.17342576384544373
step: 140, loss: 0.12760286033153534
step: 150, loss: 0.06687428057193756
step: 160, loss: 0.020797820761799812
step: 170, loss: 0.13644403219223022
step: 180, loss: 0.21192865073680878
step: 190, loss: 0.08868920058012009
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.7526881720430108, f1=0.7552083333333333,
best_f1=0.7552083333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.19522824883460999
step: 10, loss: 0.19606690108776093
step: 20, loss: 0.0900493636727333
step: 30, loss: 0.02692091092467308
step: 40, loss: 0.056619904935359955
step: 50, loss: 0.20185397565364838
step: 60, loss: 0.046434834599494934
step: 70, loss: 0.0741557627916336
step: 80, loss: 0.25372982025146484
step: 90, loss: 0.07114976644515991
step: 100, loss: 0.07570145279169083
step: 110, loss: 0.21443350613117218
step: 120, loss: 0.010249507613480091
step: 130, loss: 0.004223590716719627
step: 140, loss: 0.18980878591537476
step: 150, loss: 0.0655372366309166
step: 160, loss: 0.11820217967033386
step: 170, loss: 0.041191402822732925
step: 180, loss: 0.0994001254439354
step: 190, loss: 0.05034976452589035
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.7768595041322315, f1=0.7704918032786885,
best_f1=0.7704918032786885
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11908785253763199
step: 10, loss: 0.018808020278811455
step: 20, loss: 0.015599408186972141
step: 30, loss: 0.015446682460606098
step: 40, loss: 0.009152475744485855
step: 50, loss: 0.031096145510673523
step: 60, loss: 0.09406998008489609
step: 70, loss: 0.024921411648392677
step: 80, loss: 0.08478068560361862
step: 90, loss: 0.01021954882889986
step: 100, loss: 0.019997818395495415
step: 110, loss: 0.06099497526884079
step: 120, loss: 0.018526604399085045
step: 130, loss: 0.22661840915679932
step: 140, loss: 0.05977976322174072
step: 150, loss: 0.02657242864370346

```

```

step: 160, loss: 0.03642316162586212
step: 170, loss: 0.012314618565142155
step: 180, loss: 0.030504224821925163
step: 190, loss: 0.03048429824411869
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.8064516129032259, f1=0.7546174142480211,
best_f1=0.7546174142480211
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09640327095985413
step: 10, loss: 0.012977558188140392
step: 20, loss: 0.09113534539937973
step: 30, loss: 0.02909051440656185
step: 40, loss: 0.05221134051680565
step: 50, loss: 0.01518627256155014
step: 60, loss: 0.004292334895581007
step: 70, loss: 0.0030591259710490704
step: 80, loss: 0.021675733849406242
step: 90, loss: 0.007131052669137716
step: 100, loss: 0.04411524534225464
step: 110, loss: 0.020020976662635803
step: 120, loss: 0.005255197174847126
step: 130, loss: 0.06705117225646973
step: 140, loss: 0.005126515869051218
step: 150, loss: 0.01174501795321703
step: 160, loss: 0.010858438909053802
step: 170, loss: 0.012943967245519161
step: 180, loss: 0.020999252796173096
step: 190, loss: 0.13080716133117676
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.8169014084507042, f1=0.7731092436974789,
best_f1=0.7731092436974789
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.009059521369636059
step: 10, loss: 0.0007074947352521122
step: 20, loss: 0.041037559509277344
step: 30, loss: 0.05774233490228653
step: 40, loss: 0.10042649507522583
step: 50, loss: 0.014103448949754238
step: 60, loss: 0.006029440090060234
step: 70, loss: 0.10110461711883545
step: 80, loss: 0.12833024561405182
step: 90, loss: 0.3420431315898895
step: 100, loss: 0.06835531443357468
step: 110, loss: 0.05140329897403717
step: 120, loss: 0.014460228383541107
step: 130, loss: 0.003322309348732233
step: 140, loss: 0.001006371807307005
step: 150, loss: 0.027003442868590355
step: 160, loss: 0.003714371472597122
step: 170, loss: 0.04785007983446121
step: 180, loss: 0.01384026650339365
step: 190, loss: 0.029007649049162865
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.7816711590296496, f1=0.7595628415300547,
best_f1=0.7731092436974789
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06344648450613022
step: 10, loss: 0.017391858622431755
step: 20, loss: 0.0009929090738296509
step: 30, loss: 0.04962582141160965
step: 40, loss: 0.04584475979208946
step: 50, loss: 0.17016083002090454
step: 60, loss: 0.006682246923446655
step: 70, loss: 0.004215300548821688
step: 80, loss: 0.022383077070116997
step: 90, loss: 0.01453804038465023
step: 100, loss: 0.003934717271476984
step: 110, loss: 0.003500379389151931
step: 120, loss: 0.022911597043275833
step: 130, loss: 0.002957274205982685
step: 140, loss: 0.01596045307815075
step: 150, loss: 0.18885545432567596
step: 160, loss: 0.0030345984268933535
step: 170, loss: 0.12383998930454254
step: 180, loss: 0.008991492912173271
step: 190, loss: 0.0070610735565423965

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.7862796833773087, f1=0.7626666666666667,
best_f1=0.7731092436974789
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005071605090051889
step: 10, loss: 0.010767148807644844
step: 20, loss: 0.001365219010040164
step: 30, loss: 0.002589539624750614
step: 40, loss: 0.005936353001743555
step: 50, loss: 0.0033925778698176146
step: 60, loss: 0.020197981968522072
step: 70, loss: 0.0018688684795051813
step: 80, loss: 0.07724752277135849
step: 90, loss: 0.09128212183713913
step: 100, loss: 0.04381561651825905
step: 110, loss: 0.010669633746147156
step: 120, loss: 0.10143037885427475
step: 130, loss: 0.005424194969236851
step: 140, loss: 0.06482185423374176
step: 150, loss: 0.0008791714208200574
step: 160, loss: 0.008188802748918533
step: 170, loss: 0.0020492984913289547
step: 180, loss: 0.010219394229352474
step: 190, loss: 0.06675578653812408
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.7830687830687831, f1=0.7795698924731183,
best_f1=0.7731092436974789
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.027709180489182472
step: 10, loss: 0.0025366097688674927
step: 20, loss: 0.12707938253879547
step: 30, loss: 0.017072264105081558
step: 40, loss: 0.006658687721937895
step: 50, loss: 0.0029511067550629377
step: 60, loss: 0.0030730271246284246
step: 70, loss: 0.0018862186698243022
step: 80, loss: 0.01830686628818512
step: 90, loss: 0.00836404412984848
step: 100, loss: 0.015308951027691364
step: 110, loss: 0.0011069370666518807

```

```

step: 120, loss: 0.0030166117940098047
step: 130, loss: 0.0010935781756415963
step: 140, loss: 0.0011722269700840116
step: 150, loss: 0.017931492999196053
step: 160, loss: 0.014142248779535294
step: 170, loss: 0.02017817460000515
step: 180, loss: 0.0024903409648686647
step: 190, loss: 0.000468910438939929
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.7908163265306122, f1=0.7780678851174935,
best_f1=0.7731092436974789
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00019280942797195166
step: 10, loss: 0.018590182065963745
step: 20, loss: 0.0016137720085680485
step: 30, loss: 0.01336628943681717
step: 40, loss: 0.0005470453179441392
step: 50, loss: 0.00045066585880704224
step: 60, loss: 0.02372734621167183
step: 70, loss: 0.0008322758949361742
step: 80, loss: 0.0043061235919594765
step: 90, loss: 0.050092656165361404
step: 100, loss: 0.0032387825194746256
step: 110, loss: 0.002428625710308552
step: 120, loss: 0.01524387951940298
step: 130, loss: 0.04067125916481018
step: 140, loss: 0.0006065208581276238
step: 150, loss: 0.00023827250697650015
step: 160, loss: 0.0004200136463623494
step: 170, loss: 0.0005054877256043255
step: 180, loss: 0.03566655516624451
step: 190, loss: 0.0014333507278934121
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.7967032967032966, f1=0.7897727272727274,
best_f1=0.7731092436974789
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00651492178440094
step: 10, loss: 0.0017284287605434656
step: 20, loss: 0.0008753676083870232
step: 30, loss: 0.02258327417075634

```

```

step: 40, loss: 0.00032246430055238307
step: 50, loss: 0.0008698738529346883
step: 60, loss: 0.0006956075085327029
step: 70, loss: 0.00019542960217222571
step: 80, loss: 0.000780351459980011
step: 90, loss: 0.012186597101390362
step: 100, loss: 0.001420288230292499
step: 110, loss: 0.000849100761115551
step: 120, loss: 0.0023790462873876095
step: 130, loss: 0.0017714824061840773
step: 140, loss: 0.020491428673267365
step: 150, loss: 0.005217305850237608
step: 160, loss: 0.0047665671445429325
step: 170, loss: 0.019086457788944244
step: 180, loss: 0.06837914139032364
step: 190, loss: 0.0017253129044547677
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.7878787878787877, f1=0.775623268698061,
best_f1=0.7731092436974789
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005459211533889174
step: 10, loss: 0.00588914705440402
step: 20, loss: 0.011181874200701714
step: 30, loss: 0.018759995698928833
step: 40, loss: 0.008566446602344513
step: 50, loss: 0.0033280053175985813
step: 60, loss: 0.0005657637375406921
step: 70, loss: 0.0003967632364947349
step: 80, loss: 0.0003505840140860528
step: 90, loss: 0.0008285689982585609
step: 100, loss: 0.0005374548491090536
step: 110, loss: 0.0008670651004649699
step: 120, loss: 0.0003126774390693754
step: 130, loss: 0.0006325988797470927
step: 140, loss: 0.006013762205839157
step: 150, loss: 0.0012426909524947405
step: 160, loss: 0.10141900181770325
step: 170, loss: 0.002871062373742461
step: 180, loss: 0.00029153580544516444
step: 190, loss: 0.013403136283159256
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.8, f1=0.7878787878787877, best_f1=0.7731092436974789

```



```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0024878932163119316
step: 10, loss: 0.00509269954636693
step: 20, loss: 0.004126149695366621
step: 30, loss: 0.0005468828603625298
step: 40, loss: 0.00027174485148862004
step: 50, loss: 0.08550862967967987
step: 60, loss: 0.020453281700611115
step: 70, loss: 0.007887003943324089
step: 80, loss: 0.0007203081040643156
step: 90, loss: 0.004851267673075199
step: 100, loss: 0.11467516422271729
step: 110, loss: 0.0014301806222647429
step: 120, loss: 0.003886005375534296
step: 130, loss: 0.005997177679091692
step: 140, loss: 0.00032205297611653805
step: 150, loss: 0.0029971515759825706
step: 160, loss: 0.004323047120124102
step: 170, loss: 0.004647134803235531
step: 180, loss: 0.013916276395320892
step: 190, loss: 0.0007742183515802026
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.7989130434782609, f1=0.7845303867403315,
best_f1=0.7731092436974789
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0888720378279686
step: 10, loss: 0.003304277779534459
step: 20, loss: 0.007232019677758217
step: 30, loss: 0.00385048589669168
step: 40, loss: 0.0005345159443095326
step: 50, loss: 0.03510620817542076
step: 60, loss: 0.00049300940008834
step: 70, loss: 0.001786277280189097
step: 80, loss: 0.0010728989727795124
step: 90, loss: 0.003141851630061865
step: 100, loss: 0.0038440341595560312
step: 110, loss: 0.0006328377639874816
step: 120, loss: 0.001614136970601976
step: 130, loss: 0.0018806138541549444
step: 140, loss: 0.0012637407053261995
step: 150, loss: 0.0027745049446821213
step: 160, loss: 0.027268869802355766

```

```

step: 170, loss: 0.07077644020318985
step: 180, loss: 0.00501688988879323
step: 190, loss: 0.00044169434113427997
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.7967914438502673, f1=0.7923497267759563,
best_f1=0.7731092436974789
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002664897358044982
step: 10, loss: 0.000907589856069535
step: 20, loss: 0.006196945905685425
step: 30, loss: 0.0004578141961246729
step: 40, loss: 0.0016600966919213533
step: 50, loss: 0.0016276355599984527
step: 60, loss: 0.0016117076156660914
step: 70, loss: 0.0010782863246276975
step: 80, loss: 0.008664376102387905
step: 90, loss: 0.0031223627738654613
step: 100, loss: 0.0013719318667426705
step: 110, loss: 0.0005274050054140389
step: 120, loss: 0.0021400039549916983
step: 130, loss: 0.0012851778883486986
step: 140, loss: 0.004905689507722855
step: 150, loss: 0.0006765468860976398
step: 160, loss: 0.014226768165826797
step: 170, loss: 0.004640302155166864
step: 180, loss: 0.0008414484327659011
step: 190, loss: 0.001940999529324472
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.7956989247311829, f1=0.7857142857142857,
best_f1=0.7731092436974789
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

```

Defaults for this optimization level are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

```

Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2049it [00:05, 354.20it/s]
load_f1 = 0.6736842105263158
real_f1 = 0.6486486486486486
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:11, 389.89it/s]

```

4.2 DITTO TEXTUAL

4.2.1 Abt-Buy - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Textual/Abt-Buy \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model \
  --dk general \
  --dk product \

```

```

--summarize \
--da entry_swap \
--da attr_del

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/Abt-Buy.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
--task Textual/Abt-Buy \
--input_path input/input_small.jsonl \
--output_path output/output_small.jsonl \
--lm distilbert \
--use_gpu \
--fp16 \
--checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:

UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.

In PyTorch 1.1.0 and later, you should call them in the opposite order: ``optimizer.step()`` before ``lr_scheduler.step()``. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

```
"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
```

```
step: 0, loss: 0.8400564193725586
```

```
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
```

```
step: 10, loss: 0.22207587957382202
```

```
step: 20, loss: 0.15443266928195953
```

```
step: 30, loss: 0.24045047163963318
```

```
step: 40, loss: 0.3105899691581726
```

```
step: 50, loss: 0.3752292990684509
```

```
step: 60, loss: 0.43578752875328064
```

```
step: 70, loss: 0.3177547752857208
```

```
step: 80, loss: 0.2558150589466095
```

```
step: 90, loss: 0.4013993442058563
```

```
step: 100, loss: 0.2364388406276703
```

```
step: 110, loss: 0.1827121078968048
```

```
step: 120, loss: 0.5318783521652222
```

```
step: 130, loss: 0.42280521988868713
```

```
step: 140, loss: 0.4509902000427246
```

```
step: 150, loss: 0.06665128469467163
```

```
step: 160, loss: 0.26265326142311096
```

```
step: 170, loss: 0.1527235060930252
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

```
An input tensor was not cuda.
```

```
warnings.warn("An input tensor was not cuda.")
```

```
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
```

```
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
```

```
warnings.warn(SAVE_STATE_WARNING, UserWarning)
```

```
epoch 1: dev_f1=0.7041564792176038, f1=0.6779661016949152,
```

```
best_f1=0.6779661016949152
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

```
An input tensor was not cuda.
```

```
warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.20691712200641632
```

```
step: 10, loss: 0.08698324859142303
```

```
step: 20, loss: 0.21266911923885345
```

```
step: 30, loss: 0.19776961207389832
```

```
step: 40, loss: 0.14779454469680786
```

```
step: 50, loss: 0.16406841576099396
```

```
step: 60, loss: 0.06235500052571297
```

```
step: 70, loss: 0.22789660096168518
```

```
step: 80, loss: 0.2025974839925766
```

```
step: 90, loss: 0.1577940583229065
```

```

step: 100, loss: 0.11980094015598297
step: 110, loss: 0.1799859255552292
step: 120, loss: 0.029906081035733223
step: 130, loss: 0.04058768227696419
step: 140, loss: 0.0804884284734726
step: 150, loss: 0.15974688529968262
step: 160, loss: 0.0809665396809578
step: 170, loss: 0.1458660513162613
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.7651331719128328, f1=0.7534883720930232,
best_f1=0.7534883720930232
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.030204858630895615
step: 10, loss: 0.0651618018746376
step: 20, loss: 0.04399493336677551
step: 30, loss: 0.3015890121459961
step: 40, loss: 0.02723473124206066
step: 50, loss: 0.1536143273115158
step: 60, loss: 0.082497738301754
step: 70, loss: 0.13773564994335175
step: 80, loss: 0.06635142117738724
step: 90, loss: 0.11324919760227203
step: 100, loss: 0.054900020360946655
step: 110, loss: 0.02846217341721058
step: 120, loss: 0.02184038609266281
step: 130, loss: 0.20897284150123596
step: 140, loss: 0.025794366374611855
step: 150, loss: 0.025763090699911118
step: 160, loss: 0.24182751774787903
step: 170, loss: 0.1434415578842163
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.8041237113402061, f1=0.7969543147208121,
best_f1=0.7969543147208121
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1161646768450737
step: 10, loss: 0.10252950340509415
step: 20, loss: 0.021347224712371826
step: 30, loss: 0.10766114294528961
step: 40, loss: 0.07968784868717194
step: 50, loss: 0.009482195600867271
step: 60, loss: 0.1928231567144394
step: 70, loss: 0.025049123913049698
step: 80, loss: 0.06912383437156677
step: 90, loss: 0.04988100379705429
step: 100, loss: 0.008546348661184311
step: 110, loss: 0.038388971239328384
step: 120, loss: 0.13183057308197021
step: 130, loss: 0.0803680419921875
step: 140, loss: 0.011935249902307987
step: 150, loss: 0.02312350645661354
step: 160, loss: 0.05657313019037247
step: 170, loss: 0.014211773872375488
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.8132992327365729, f1=0.7980049875311721,
best_f1=0.7980049875311721
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09276016056537628
step: 10, loss: 0.08407849073410034
step: 20, loss: 0.13174690306186676
step: 30, loss: 0.09577293694019318
step: 40, loss: 0.02650456503033638
step: 50, loss: 0.03787732869386673
step: 60, loss: 0.0020629088394343853
step: 70, loss: 0.00886794924736023
step: 80, loss: 0.005943183787167072
step: 90, loss: 0.02451349049806595
step: 100, loss: 0.03407187759876251
step: 110, loss: 0.10766824334859848
step: 120, loss: 0.035791754722595215
step: 130, loss: 0.00825521256774664
step: 140, loss: 0.029508240520954132
step: 150, loss: 0.01362450048327446

```

```

step: 160, loss: 0.002552703255787492
step: 170, loss: 0.08492392301559448
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 5: dev_f1=0.8211586901763224, f1=0.8089330024813897,
best_f1=0.8089330024813897
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01926000416278839
step: 10, loss: 0.08554121851921082
step: 20, loss: 0.04357939958572388
step: 30, loss: 0.011074936017394066
step: 40, loss: 0.13480781018733978
step: 50, loss: 0.011884858831763268
step: 60, loss: 0.008757967501878738
step: 70, loss: 0.016134334728121758
step: 80, loss: 0.12647876143455505
step: 90, loss: 0.06578965485095978
step: 100, loss: 0.024539029225707054
step: 110, loss: 0.00917427334934473
step: 120, loss: 0.059725578874349594
step: 130, loss: 0.17996762692928314
step: 140, loss: 0.00958999153226614
step: 150, loss: 0.19922515749931335
step: 160, loss: 0.06945236772298813
step: 170, loss: 0.005673231091350317
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.8192771084337348, f1=0.8111888111888111,
best_f1=0.8089330024813897
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005987490992993116
step: 10, loss: 0.00284404749982059
step: 20, loss: 0.053730666637420654
step: 30, loss: 0.0029217672999948263
step: 40, loss: 0.013150317594408989
step: 50, loss: 0.0007219689432531595
step: 60, loss: 0.25668883323669434
step: 70, loss: 0.019297420978546143

```



```

step: 80, loss: 0.027275221422314644
step: 90, loss: 0.04447167366743088
step: 100, loss: 0.008960632607340813
step: 110, loss: 0.006452437490224838
step: 120, loss: 0.1657479703426361
step: 130, loss: 0.23536109924316406
step: 140, loss: 0.01107864361256361
step: 150, loss: 0.07188185304403305
step: 160, loss: 0.010052149184048176
step: 170, loss: 0.18052791059017181
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.8177339901477833, f1=0.8066825775656326,
best_f1=0.8089330024813897
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02815985679626465
step: 10, loss: 0.03833477944135666
step: 20, loss: 0.004838207736611366
step: 30, loss: 0.004068007227033377
step: 40, loss: 0.0045913029462099075
step: 50, loss: 0.0610274001955986
step: 60, loss: 0.0018362341215834022
step: 70, loss: 0.061006322503089905
step: 80, loss: 0.025982901453971863
step: 90, loss: 0.045191362500190735
step: 100, loss: 0.0038753061089664698
step: 110, loss: 0.015134632587432861
step: 120, loss: 0.0033749411813914776
step: 130, loss: 0.019746368750929832
step: 140, loss: 0.04495886713266373
step: 150, loss: 0.12079337984323502
step: 160, loss: 0.01402085367590189
step: 170, loss: 0.03296497464179993
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 8: dev_f1=0.8317307692307692, f1=0.8243559718969554,
best_f1=0.8243559718969554
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.021069597452878952
step: 10, loss: 0.06121228262782097
step: 20, loss: 0.0016735197277739644
step: 30, loss: 0.020136758685112
step: 40, loss: 0.027642445638775826
step: 50, loss: 0.023388151079416275
step: 60, loss: 0.005198479164391756
step: 70, loss: 0.02157902903854847
step: 80, loss: 0.06910455226898193
step: 90, loss: 0.009390673600137234
step: 100, loss: 0.03853137791156769
step: 110, loss: 0.0022152008023113012
step: 120, loss: 0.001303485594689846
step: 130, loss: 0.006750243715941906
step: 140, loss: 0.06986134499311447
step: 150, loss: 0.01729436218738556
step: 160, loss: 0.0054868427105247974
step: 170, loss: 0.023383162915706635
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 9: dev_f1=0.818627450980392, f1=0.7822014051522249,
best_f1=0.8243559718969554
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.016105877235531807
step: 10, loss: 0.03185895085334778
step: 20, loss: 0.0019512349972501397
step: 30, loss: 0.0006798954564146698
step: 40, loss: 0.03184647858142853
step: 50, loss: 0.028805667534470558
step: 60, loss: 0.003635202534496784
step: 70, loss: 0.0009355733636766672
step: 80, loss: 0.01688554883003235
step: 90, loss: 0.1376059204339981
step: 100, loss: 0.00028276978991925716
step: 110, loss: 0.10032712668180466
step: 120, loss: 0.011260212399065495
step: 130, loss: 0.031599145382642746
step: 140, loss: 0.0024597621522843838
step: 150, loss: 0.004734716843813658
step: 160, loss: 0.01569889858365059
step: 170, loss: 0.02450701780617237
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.8260869565217391, f1=0.8056872037914692,

```

```

best_f1=0.8243559718969554
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00033787317806854844
step: 10, loss: 0.01129030529409647
step: 20, loss: 0.017516957595944405
step: 30, loss: 0.002815989078953862
step: 40, loss: 0.009535145945847034
step: 50, loss: 0.0006538084708154202
step: 60, loss: 0.012737479992210865
step: 70, loss: 0.005137849133461714
step: 80, loss: 0.0013297456316649914
step: 90, loss: 0.003692115191370249
step: 100, loss: 0.0004995661438442767
step: 110, loss: 0.003782588057219982
step: 120, loss: 0.027805084362626076
step: 130, loss: 0.012913635931909084
step: 140, loss: 0.026303038001060486
step: 150, loss: 0.0022959152702242136
step: 160, loss: 0.036622632294893265
step: 170, loss: 0.051735877990722656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.8213333333333334, f1=0.8031088082901555,
best_f1=0.8243559718969554
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014395908452570438
step: 10, loss: 0.018615374341607094
step: 20, loss: 0.030223729088902473
step: 30, loss: 0.0013919929042458534
step: 40, loss: 0.0014494413044303656
step: 50, loss: 0.009218990802764893
step: 60, loss: 0.24648843705654144
step: 70, loss: 0.021929318085312843
step: 80, loss: 0.021295025944709778
step: 90, loss: 0.00033526637707836926
step: 100, loss: 0.0008756747702136636
step: 110, loss: 0.0006921145250089467
step: 120, loss: 0.05141206458210945
step: 130, loss: 0.01000931952148676
step: 140, loss: 0.0007487904513254762
step: 150, loss: 0.007223881781101227
step: 160, loss: 0.00916603859513998
step: 170, loss: 0.0034273152705281973

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.8211586901763224, f1=0.8164251207729469,
best_f1=0.8243559718969554
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0028039016760885715
step: 10, loss: 0.07157805562019348
step: 20, loss: 0.010728778317570686
step: 30, loss: 0.0037238621152937412
step: 40, loss: 0.2142646163702011
step: 50, loss: 0.005577957723289728
step: 60, loss: 0.020650167018175125
step: 70, loss: 0.059106457978487015
step: 80, loss: 0.0029898248612880707
step: 90, loss: 0.0008249891106970608
step: 100, loss: 0.0022853135596960783
step: 110, loss: 0.013710146769881248
step: 120, loss: 0.0009280518279410899
step: 130, loss: 0.01061225589364767
step: 140, loss: 0.0006231775041669607
step: 150, loss: 0.0016824222402647138
step: 160, loss: 0.0012678292114287615
step: 170, loss: 0.003709131618961692
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.8309178743961353, f1=0.8142857142857144,
best_f1=0.8243559718969554
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005433216691017151
step: 10, loss: 0.0007231730269268155
step: 20, loss: 0.014322380535304546
step: 30, loss: 0.001104296068660915
step: 40, loss: 0.0012680444633588195
step: 50, loss: 0.04029029607772827
step: 60, loss: 0.00100386468693614
step: 70, loss: 0.011266111396253109
step: 80, loss: 0.004191101528704166
step: 90, loss: 0.0012582764029502869
step: 100, loss: 0.014661308377981186
step: 110, loss: 0.002196842571720481
step: 120, loss: 0.003169685136526823
step: 130, loss: 0.0021616427693516016

```

```

step: 140, loss: 0.0012318630469962955
step: 150, loss: 0.003631644882261753
step: 160, loss: 0.05952712148427963
step: 170, loss: 0.0006578747415915132
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 14: dev_f1=0.832535885167464, f1=0.815165876777251,
best_f1=0.815165876777251
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006864107213914394
step: 10, loss: 0.0016134475590661168
step: 20, loss: 0.012423979118466377
step: 30, loss: 0.030361726880073547
step: 40, loss: 0.0009578106692060828
step: 50, loss: 0.0035184472799301147
step: 60, loss: 0.000587272341363132
step: 70, loss: 0.0013198790838941932
step: 80, loss: 0.000814635946881026
step: 90, loss: 0.0006764794234186411
step: 100, loss: 0.000603383406996727
step: 110, loss: 0.0010619156528264284
step: 120, loss: 0.007689686492085457
step: 130, loss: 0.0020624944008886814
step: 140, loss: 0.002787893870845437
step: 150, loss: 0.05075807869434357
step: 160, loss: 0.0004611655022017658
step: 170, loss: 0.007205377332866192
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 15: dev_f1=0.8337236533957846, f1=0.8045977011494253,
best_f1=0.8045977011494253
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

```

Defaults for this optimization level are:

```

enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
1916it [00:04, 443.99it/s]
load_f1 = 0.41666666666666666
real_f1 = 0.3902439024390244
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:11, 399.12it/s]

```

4.3 DITTO DIRTY

4.3.1 DBLP-ACM - Running the matcher

```

[ ]: # Step 1:
      # train the wdc model with the save_model flag on
      # it will produce two files *_dev.pt and *_test.pt which are the
      # best checkpoints on the dev set and the test set.
      !CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
        --task Dirty/DBLP-ACM \
        --batch_size 32 \
        --max_len 256 \
        --lr 3e-5 \

```

```

--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \
--dk product \
--summarize \
--da entry_swap \
--da attr_del

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-ACM.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Dirty/DBLP-ACM \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic

```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.

```

Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8060435056686401
step: 10, loss: 0.4649621546268463
step: 20, loss: 0.5593419671058655
step: 30, loss: 0.3641819655895233
step: 40, loss: 0.15901057422161102
step: 50, loss: 0.12626121938228607
step: 60, loss: 0.06937345862388611
step: 70, loss: 0.09998650848865509
step: 80, loss: 0.2016032636165619
step: 90, loss: 0.03663868457078934
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 100, loss: 0.1742790937423706
step: 110, loss: 0.02529674954712391
step: 120, loss: 0.06127670034766197
step: 130, loss: 0.01821926422417164
step: 140, loss: 0.041105855256319046
step: 150, loss: 0.13470324873924255
step: 160, loss: 0.18310537934303284
step: 170, loss: 0.020983999595046043
step: 180, loss: 0.022825421765446663
step: 190, loss: 0.06140230596065521
step: 200, loss: 0.005038405302911997
step: 210, loss: 0.0077169849537312984
step: 220, loss: 0.014150136150419712
step: 230, loss: 0.027767635881900787
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9525909592061743, f1=0.9588431590656283,
best_f1=0.9588431590656283

```



```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.20462524890899658
step: 10, loss: 0.005089688114821911
step: 20, loss: 0.0027797489892691374
step: 30, loss: 0.0041950661689043045
step: 40, loss: 0.009745090268552303
step: 50, loss: 0.023045659065246582
step: 60, loss: 0.008877050131559372
step: 70, loss: 0.04702022671699524
step: 80, loss: 0.02345380373299122
step: 90, loss: 0.02501082234084606
step: 100, loss: 0.13442407548427582
step: 110, loss: 0.09393353760242462
step: 120, loss: 0.0013027724344283342
step: 130, loss: 0.025752900168299675
step: 140, loss: 0.17209339141845703
step: 150, loss: 0.01691717654466629
step: 160, loss: 0.01032341830432415
step: 170, loss: 0.02511323057115078
step: 180, loss: 0.019282560795545578
step: 190, loss: 0.08222134411334991
step: 200, loss: 0.007332527078688145
step: 210, loss: 0.1803465038537979
step: 220, loss: 0.0012321284739300609
step: 230, loss: 0.02127225324511528
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 2: dev_f1=0.9647577092511014, f1=0.9623893805309734,
best_f1=0.9623893805309734
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11970970034599304
step: 10, loss: 0.03245539590716362
step: 20, loss: 0.004260519985109568
step: 30, loss: 0.1454905867576599
step: 40, loss: 0.035862475633621216
step: 50, loss: 0.014255084097385406
step: 60, loss: 0.03882655128836632
step: 70, loss: 0.002118546748533845
step: 80, loss: 0.0012782533885911107

```

```

step: 90, loss: 0.004852375481277704
step: 100, loss: 0.0018233929295092821
step: 110, loss: 0.027822919189929962
step: 120, loss: 0.01613042689859867
step: 130, loss: 0.0023635674733668566
step: 140, loss: 0.024836266413331032
step: 150, loss: 0.013804781250655651
step: 160, loss: 0.016323797404766083
step: 170, loss: 0.04155227541923523
step: 180, loss: 0.0037525512743741274
step: 190, loss: 0.008865300565958023
step: 200, loss: 0.004241872578859329
step: 210, loss: 0.010919378139078617
step: 220, loss: 0.019261270761489868
step: 230, loss: 0.06428734958171844
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.961883408071749, f1=0.9632925472747497,
best_f1=0.9623893805309734
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.027779797092080116
step: 10, loss: 0.10713663697242737
step: 20, loss: 0.0043166750110685825
step: 30, loss: 0.0017210893565788865
step: 40, loss: 0.008872454054653645
step: 50, loss: 0.0031039579771459103
step: 60, loss: 0.0029847226105630398
step: 70, loss: 0.08384761959314346
step: 80, loss: 0.09948348253965378
step: 90, loss: 0.022670866921544075
step: 100, loss: 0.0013539890060201287
step: 110, loss: 0.0988805890083313
step: 120, loss: 0.13770511746406555
step: 130, loss: 0.03890803083777428
step: 140, loss: 0.0008397332858294249
step: 150, loss: 0.018185624852776527
step: 160, loss: 0.007631795480847359
step: 170, loss: 0.03834239020943642
step: 180, loss: 0.11073742061853409
step: 190, loss: 0.0036099811550229788
step: 200, loss: 0.0018784446874633431
step: 210, loss: 0.0030324277468025684
step: 220, loss: 0.0043501583859324455
step: 230, loss: 0.0007325134356506169
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.9700996677740864, f1=0.9654403567447045,
best_f1=0.9654403567447045
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006066776113584638
step: 10, loss: 0.0004485116805881262
step: 20, loss: 0.00040742941200733185
step: 30, loss: 0.0015429715858772397
step: 40, loss: 0.0004642456478904933
step: 50, loss: 0.00022313318913802505
step: 60, loss: 0.0005941791459918022
step: 70, loss: 0.00020031121675856411
step: 80, loss: 0.000614230870269239
step: 90, loss: 0.0013824517372995615
step: 100, loss: 0.01805397868156433
step: 110, loss: 0.0020036869682371616
step: 120, loss: 0.11666261404752731
step: 130, loss: 0.006083231884986162
step: 140, loss: 0.0025620285887271166
step: 150, loss: 0.0003054978442378342
step: 160, loss: 0.10674639791250229
step: 170, loss: 0.04179494455456734
step: 180, loss: 0.0019048925023525953
step: 190, loss: 0.005395404528826475
step: 200, loss: 0.004586872179061174
step: 210, loss: 0.003274050774052739
step: 220, loss: 0.00956481508910656
step: 230, loss: 0.0018945086048915982
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.96875, f1=0.9709821428571428, best_f1=0.9654403567447045
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0018605014774948359
step: 10, loss: 0.00023688121291343123
step: 20, loss: 0.0025524392258375883
step: 30, loss: 0.002229099627584219
step: 40, loss: 0.0029433921445161104
step: 50, loss: 0.1407838761806488

```

```

step: 60, loss: 0.004017238970845938
step: 70, loss: 0.0009478969150222838
step: 80, loss: 0.0005242625484243035
step: 90, loss: 0.00041387745295651257
step: 100, loss: 0.0014470937894657254
step: 110, loss: 0.045218247920274734
step: 120, loss: 0.006787079852074385
step: 130, loss: 0.002587519586086273
step: 140, loss: 0.0008926399750635028
step: 150, loss: 0.0001980933448066935
step: 160, loss: 0.002996747149154544
step: 170, loss: 0.05820673331618309
step: 180, loss: 0.0005904090357944369
step: 190, loss: 0.014808627776801586
step: 200, loss: 0.0042694634757936
step: 210, loss: 0.00118590600322932
step: 220, loss: 0.0005597811541520059
step: 230, loss: 0.002705564023926854
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 6: dev_f1=0.9653631284916202, f1=0.9720044792833147,
best_f1=0.9654403567447045
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02608376368880272
step: 10, loss: 0.006701814476400614
step: 20, loss: 0.026298172771930695
step: 30, loss: 0.0011822268133983016
step: 40, loss: 0.0015630520647391677
step: 50, loss: 0.00016168215370271355
step: 60, loss: 0.03058740496635437
step: 70, loss: 0.014007366262376308
step: 80, loss: 0.0029242485761642456
step: 90, loss: 0.004762009717524052
step: 100, loss: 0.0010887716198340058
step: 110, loss: 0.00917089730501175
step: 120, loss: 0.039994750171899796
step: 130, loss: 0.00037411166704259813
step: 140, loss: 0.00021599857427645475
step: 150, loss: 0.001252425485290587
step: 160, loss: 0.0006572767742909491
step: 170, loss: 0.0011028636945411563
step: 180, loss: 0.00027374981436878443
step: 190, loss: 0.027782723307609558
step: 200, loss: 0.06388110667467117
step: 210, loss: 0.00021483682212419808

```

```

step: 220, loss: 0.006234504282474518
step: 230, loss: 0.08703965693712234
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.9731543624161074, f1=0.9688888888888889,
best_f1=0.9688888888888889
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.023342054337263107
step: 10, loss: 0.08888690173625946
step: 20, loss: 0.0006706180865876377
step: 30, loss: 0.0008735884330235422
step: 40, loss: 0.0013603721745312214
step: 50, loss: 0.0003762318810913712
step: 60, loss: 0.0004265549941919744
step: 70, loss: 0.0019514142768457532
step: 80, loss: 0.019281191751360893
step: 90, loss: 9.953036351362243e-05
step: 100, loss: 0.0001472238072892651
step: 110, loss: 0.0010774885304272175
step: 120, loss: 0.011985392309725285
step: 130, loss: 0.10983981937170029
step: 140, loss: 0.0006243474199436605
step: 150, loss: 0.0006465689511969686
step: 160, loss: 0.0038540393579751253
step: 170, loss: 0.018326926976442337
step: 180, loss: 0.0022785465698689222
step: 190, loss: 0.0003262062091380358
step: 200, loss: 0.04129008203744888
step: 210, loss: 9.559840691508725e-05
step: 220, loss: 0.0003434153040871024
step: 230, loss: 0.029483327642083168
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 8: dev_f1=0.9743589743589743, f1=0.9689578713968958,
best_f1=0.9689578713968958
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0010171226458624005
step: 10, loss: 0.0012551890686154366
step: 20, loss: 0.0013409991515800357
step: 30, loss: 0.0009987469529733062
step: 40, loss: 0.0004716886905953288
step: 50, loss: 0.0006920926971361041
step: 60, loss: 0.0005437738727778196
step: 70, loss: 0.000369396002497524
step: 80, loss: 0.03865745663642883
step: 90, loss: 0.007691049017012119
step: 100, loss: 0.006349234376102686
step: 110, loss: 0.030861053615808487
step: 120, loss: 0.06891977041959763
step: 130, loss: 0.00010969777940772474
step: 140, loss: 0.14248156547546387
step: 150, loss: 5.600614531431347e-05
step: 160, loss: 0.0005584629252552986
step: 170, loss: 0.00041576995863579214
step: 180, loss: 0.011222777888178825
step: 190, loss: 0.007991704158484936
step: 200, loss: 0.00013817677972838283
step: 210, loss: 0.0021875689271837473
step: 220, loss: 0.03531939163804054
step: 230, loss: 0.03769758343696594
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
  warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 9: dev_f1=0.9755011135857461, f1=0.966740576496674,
best_f1=0.966740576496674
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004127908905502409
step: 10, loss: 0.0001038451591739431
step: 20, loss: 0.001191339106298983
step: 30, loss: 0.006603084970265627
step: 40, loss: 0.0009559436584822834
step: 50, loss: 0.0004687148320954293
step: 60, loss: 0.06869415193796158
step: 70, loss: 0.0021949352230876684
step: 80, loss: 0.0054915789514780045
step: 90, loss: 0.0015964360209181905

```

```

step: 100, loss: 0.00022192044707480818
step: 110, loss: 0.006232540123164654
step: 120, loss: 0.005447537172585726
step: 130, loss: 0.06334138661623001
step: 140, loss: 0.011871126480400562
step: 150, loss: 0.0014615317340940237
step: 160, loss: 0.0008027954027056694
step: 170, loss: 0.005874308291822672
step: 180, loss: 0.00029291672399267554
step: 190, loss: 0.0005369147984310985
step: 200, loss: 0.0012939526932314038
step: 210, loss: 0.00015038686979096383
step: 220, loss: 0.004746683407574892
step: 230, loss: 0.04243793711066246
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9730337078651685, f1=0.9743016759776536,
best_f1=0.966740576496674
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002295876038260758
step: 10, loss: 0.0016318836715072393
step: 20, loss: 4.6720917453058064e-05
step: 30, loss: 0.0005500568076968193
step: 40, loss: 0.00834647286683321
step: 50, loss: 0.0025716449599713087
step: 60, loss: 0.00012375708320178092
step: 70, loss: 0.00043951295083388686
step: 80, loss: 0.0002634590200614184
step: 90, loss: 0.0005561891011893749
step: 100, loss: 0.00012460621655918658
step: 110, loss: 0.0017449667211622
step: 120, loss: 0.02198108844459057
step: 130, loss: 4.24214122176636e-05
step: 140, loss: 5.65743976039812e-05
step: 150, loss: 6.16771139902994e-05
step: 160, loss: 0.0009850026108324528
step: 170, loss: 0.02096819132566452
step: 180, loss: 0.0014971133787184954
step: 190, loss: 0.0005448405863717198
step: 200, loss: 0.0004972905153408647
step: 210, loss: 6.711974856443703e-05
step: 220, loss: 6.125497748143971e-05
step: 230, loss: 0.007884415797889233
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
epoch 11: dev_f1=0.9730941704035874, f1=0.9688195991091313,
best_f1=0.966740576496674
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00042733256123028696
step: 10, loss: 0.0005171544034965336
step: 20, loss: 4.27120212407317e-05
step: 30, loss: 0.0005079364636912942
step: 40, loss: 7.851330883568153e-05
step: 50, loss: 0.00016305327881127596
step: 60, loss: 0.01372593268752098
step: 70, loss: 0.00038545927964150906
step: 80, loss: 7.392960833385587e-05
step: 90, loss: 0.00012275906919967383
step: 100, loss: 0.0006368983304128051
step: 110, loss: 0.0003881402953993529
step: 120, loss: 0.00025635745259933174
step: 130, loss: 0.1491529792547226
step: 140, loss: 0.0005342208896763623
step: 150, loss: 0.0048055499792099
step: 160, loss: 0.0019406330538913608
step: 170, loss: 0.0010544847464188933
step: 180, loss: 0.0006601495551876724
step: 190, loss: 7.207204907899722e-05
step: 200, loss: 0.004517221823334694
step: 210, loss: 0.0001271270157303661
step: 220, loss: 0.039455343037843704
step: 230, loss: 7.058588380459696e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9733333333333333, f1=0.9636963696369637,
best_f1=0.966740576496674
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003902795142494142
step: 10, loss: 0.00028695102082565427
step: 20, loss: 8.399055514018983e-05
step: 30, loss: 0.02487126551568508
step: 40, loss: 0.0026378154288977385
step: 50, loss: 0.00047356021241284907
step: 60, loss: 0.0001131405369960703
step: 70, loss: 7.675933738937601e-05
step: 80, loss: 7.777697464916855e-05
step: 90, loss: 6.825125456089154e-05

```



```

step: 100, loss: 8.194996189558879e-05
step: 110, loss: 9.072816465049982e-05
step: 120, loss: 0.00022015006106812507
step: 130, loss: 0.00012397054524626583
step: 140, loss: 0.0002692423586267978
step: 150, loss: 0.0012867714976891875
step: 160, loss: 0.0002041216939687729
step: 170, loss: 0.0007937942864373326
step: 180, loss: 0.008018224500119686
step: 190, loss: 4.9834608944365755e-05
step: 200, loss: 0.00015342305414378643
step: 210, loss: 0.00027903946465812624
step: 220, loss: 0.00039191197720356286
step: 230, loss: 6.748089072061703e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 13: dev_f1=0.9755555555555556, f1=0.9668141592920354,
best_f1=0.9668141592920354
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.2120230975560844e-05
step: 10, loss: 3.565755105228163e-05
step: 20, loss: 0.01847977191209793
step: 30, loss: 0.02118551731109619
step: 40, loss: 7.392986299237236e-05
step: 50, loss: 0.0006925773923285306
step: 60, loss: 0.0033795014023780823
step: 70, loss: 0.00010918008047156036
step: 80, loss: 7.870794070186093e-05
step: 90, loss: 0.0005064652650617063
step: 100, loss: 0.00362472515553236
step: 110, loss: 0.00013998668873682618
step: 120, loss: 0.00019118882482871413
step: 130, loss: 7.25041827536188e-05
step: 140, loss: 0.0027065924368798733
step: 150, loss: 7.128226570785046e-05
step: 160, loss: 4.6245775592979044e-05
step: 170, loss: 0.00015543143672402948
step: 180, loss: 7.037614705041051e-05
step: 190, loss: 0.0015501868911087513
step: 200, loss: 5.3563242545351386e-05
step: 210, loss: 0.015924222767353058

```

```

step: 220, loss: 9.129498357651755e-05
step: 230, loss: 6.505691999336705e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9743589743589743, f1=0.9689578713968958,
best_f1=0.9668141592920354
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00012790238542947918
step: 10, loss: 0.0001117055508075282
step: 20, loss: 4.0131359128281474e-05
step: 30, loss: 0.00021614923025481403
step: 40, loss: 0.00010722403385443613
step: 50, loss: 0.0004548634751699865
step: 60, loss: 4.9838123231893405e-05
step: 70, loss: 9.015484829433262e-05
step: 80, loss: 0.004000952001661062
step: 90, loss: 4.6938897867221385e-05
step: 100, loss: 7.267908222274855e-05
step: 110, loss: 0.00013702583964914083
step: 120, loss: 7.876075687818229e-05
step: 130, loss: 6.724072591168806e-05
step: 140, loss: 0.025145303457975388
step: 150, loss: 0.0002642463077791035
step: 160, loss: 0.00964293535798788
step: 170, loss: 5.6706594477873296e-05
step: 180, loss: 7.639179966645315e-05
step: 190, loss: 0.025299139320850372
step: 200, loss: 0.00010751641821116209
step: 210, loss: 0.029620332643389702
step: 220, loss: 0.0012562115443870425
step: 230, loss: 5.686036456609145e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.9731543624161074, f1=0.9699666295884317,
best_f1=0.9668141592920354
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

```

Defaults for this optimization level are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False

```

```

keep_batchnorm_fp32      : True
master_weights           : True
loss_scale                : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                   : True
opt_level                 : 02
cast_model_type           : torch.float16
patch_torch_functions     : False
keep_batchnorm_fp32      : True
master_weights           : True
loss_scale                : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2473it [00:08, 292.03it/s]
load_f1 = 0.9730337078651685
real_f1 = 0.9719416386083053
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:10, 401.85it/s]

```

4.3.2 DBLP-GoogleScholar - Running the matcher

```

[ ]: # Step 1:
# train the wdc model with the save_model flag on
# it will produce two files *_dev.pt and *_test.pt which are the
# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
--task Dirty/DBLP-GoogleScholar \
--batch_size 32 \
--max_len 256 \
--lr 3e-5 \
--n_epochs 15 \
--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \

```

```

--dk product \
--summarize \
--da entry_swap \
--da attr_del

# Step 2: pick a checkpoint, rename it to wdc_all_small.pt
!mv *_dev.pt checkpoints/DBLP-GoogleScholar/.pt

# Step 3: run the matcher
!CUDA_VISIBLE_DEVICES=0 python matcher.py \
  --task Dirty/DBLP-GoogleScholar \
  --input_path input/input_small.jsonl \
  --output_path output/output_small.jsonl \
  --lm distilbert \
  --use_gpu \
  --fp16 \
  --checkpoint_path checkpoints/

```

Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Processing user overrides (additional kwargs that are not None)...

After processing overrides, optimization options are:

```

enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic

```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:

UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`. In PyTorch 1.1.0 and later, you should call them in the opposite order: `optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result in PyTorch skipping the first value of the learning rate schedule. See more details at <https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>

"<https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate>", UserWarning)

```
step: 0, loss: 0.7914621233940125
step: 10, loss: 0.4164191484451294
step: 20, loss: 0.4918142259120941
step: 30, loss: 0.38577455282211304
step: 40, loss: 0.25321653485298157
step: 50, loss: 0.1522245705127716
step: 60, loss: 0.18239888548851013
step: 70, loss: 0.25636711716651917
step: 80, loss: 0.13266079127788544
step: 90, loss: 0.05711810663342476
step: 100, loss: 0.38821396231651306
step: 110, loss: 0.085743747651577
step: 120, loss: 0.06549456715583801
step: 130, loss: 0.045268211513757706
step: 140, loss: 0.145346000790596
step: 150, loss: 0.03808769956231117
step: 160, loss: 0.20347224175930023
step: 170, loss: 0.11873462051153183
step: 180, loss: 0.1171247810125351
step: 190, loss: 0.041329555213451385
step: 200, loss: 0.08625639230012894
step: 210, loss: 0.09460977464914322
step: 220, loss: 0.11865735054016113
step: 230, loss: 0.11478107422590256
step: 240, loss: 0.16280819475650787
step: 250, loss: 0.05562210828065872
step: 260, loss: 0.17359673976898193
step: 270, loss: 0.014163419604301453
step: 280, loss: 0.1211032122373581
step: 290, loss: 0.1086181104183197
step: 300, loss: 0.08920905739068985
step: 310, loss: 0.0344085693359375
step: 320, loss: 0.04918926581740379
step: 330, loss: 0.0835706815123558
step: 340, loss: 0.24853944778442383
step: 350, loss: 0.1364484280347824
step: 360, loss: 0.08744063228368759
step: 370, loss: 0.11108870059251785
step: 380, loss: 0.1876598298549652
step: 390, loss: 0.02370363660156727
```

```

step: 400, loss: 0.007879288867115974
step: 410, loss: 0.029731150716543198
step: 420, loss: 0.005937084089964628
step: 430, loss: 0.04553665220737457
step: 440, loss: 0.059356771409511566
step: 450, loss: 0.027699517086148262
step: 460, loss: 0.16414916515350342
step: 470, loss: 0.1257115602493286
step: 480, loss: 0.26404163241386414
step: 490, loss: 0.019299767911434174
step: 500, loss: 0.01070642564445734
step: 510, loss: 0.06108696758747101
step: 520, loss: 0.0658596083521843
step: 530, loss: 0.06067746505141258
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 1: dev_f1=0.9362292051756008, f1=0.9324137931034482,
best_f1=0.9324137931034482
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11470160633325577
step: 10, loss: 0.14754199981689453
step: 20, loss: 0.0618765652179718
step: 30, loss: 0.035941705107688904
step: 40, loss: 0.003040025942027569
step: 50, loss: 0.10866311192512512
step: 60, loss: 0.11936118453741074
step: 70, loss: 0.29524287581443787
step: 80, loss: 0.020442137494683266
step: 90, loss: 0.018760506063699722
step: 100, loss: 0.5033866167068481
step: 110, loss: 0.05511777848005295
step: 120, loss: 0.058951087296009064
step: 130, loss: 0.017749985679984093
step: 140, loss: 0.01931268349289894
step: 150, loss: 0.04022419452667236
step: 160, loss: 0.015160849317908287
step: 170, loss: 0.1459398716688156
step: 180, loss: 0.014214636757969856
step: 190, loss: 0.03263459354639053
step: 200, loss: 0.007652048021554947
step: 210, loss: 0.006932580843567848

```

```
step: 220, loss: 0.15051233768463135
step: 230, loss: 0.009271996095776558
step: 240, loss: 0.12724395096302032
step: 250, loss: 0.027514679357409477
step: 260, loss: 0.019320394843816757
step: 270, loss: 0.17445124685764313
step: 280, loss: 0.2545962929725647
step: 290, loss: 0.06763333082199097
step: 300, loss: 0.06226417049765587
step: 310, loss: 0.06683547794818878
step: 320, loss: 0.0843050628900528
step: 330, loss: 0.059308987110853195
step: 340, loss: 0.008607210591435432
step: 350, loss: 0.06502058357000351
step: 360, loss: 0.015769729390740395
step: 370, loss: 0.005745554342865944
step: 380, loss: 0.05866999551653862
step: 390, loss: 0.041850652545690536
step: 400, loss: 0.04837203770875931
step: 410, loss: 0.0010352026438340545
step: 420, loss: 0.06894595175981522
step: 430, loss: 0.012918462045490742
step: 440, loss: 0.011996370740234852
step: 450, loss: 0.02619503252208233
step: 460, loss: 0.18949094414710999
step: 470, loss: 0.04171070083975792
step: 480, loss: 0.18239617347717285
step: 490, loss: 0.02970128133893013
step: 500, loss: 0.050936948508024216
step: 510, loss: 0.03057747334241867
step: 520, loss: 0.10633505880832672
step: 530, loss: 0.153197780251503
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 2: dev_f1=0.9349930843706776, f1=0.9316200091785223,
best_f1=0.9324137931034482
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008035737089812756
step: 10, loss: 0.08490914106369019
step: 20, loss: 0.15865245461463928
step: 30, loss: 0.1336955726146698
step: 40, loss: 0.003380183130502701
step: 50, loss: 0.026112254709005356
step: 60, loss: 0.005695842206478119
step: 70, loss: 0.012103459797799587
```

step: 80, loss: 0.008068732917308807
step: 90, loss: 0.08768428862094879
step: 100, loss: 0.0755186378955841
step: 110, loss: 0.017055600881576538
step: 120, loss: 0.059567008167505264
step: 130, loss: 0.017373189330101013
step: 140, loss: 0.06511716544628143
step: 150, loss: 0.0030526320915669203
step: 160, loss: 0.01940257102251053
step: 170, loss: 0.009477701038122177
step: 180, loss: 0.007731516379863024
step: 190, loss: 0.016352543607354164
step: 200, loss: 0.014214995317161083
step: 210, loss: 0.06372714042663574
step: 220, loss: 0.01738031767308712
step: 230, loss: 0.011233903467655182
step: 240, loss: 0.01775273121893406
step: 250, loss: 0.0038984583225101233
step: 260, loss: 0.005856987088918686
step: 270, loss: 0.008970146998763084
step: 280, loss: 0.010552924126386642
step: 290, loss: 0.05861354619264603
step: 300, loss: 0.21331025660037994
step: 310, loss: 0.028419731184840202
step: 320, loss: 0.12755873799324036
step: 330, loss: 0.001406629104167223
step: 340, loss: 0.005088559351861477
step: 350, loss: 0.05994925647974014
step: 360, loss: 0.016704194247722626
step: 370, loss: 0.023462839424610138
step: 380, loss: 0.0908658355474472
step: 390, loss: 0.026016965508461
step: 400, loss: 0.022985735908150673
step: 410, loss: 0.012617512606084347
step: 420, loss: 0.07406371831893921
step: 430, loss: 0.011262305080890656
step: 440, loss: 0.06018288806080818
step: 450, loss: 0.07253911346197128
step: 460, loss: 0.08966848254203796
step: 470, loss: 0.0018326970748603344
step: 480, loss: 0.06471730768680573
step: 490, loss: 0.01913972571492195
step: 500, loss: 0.023625757545232773
step: 510, loss: 0.005133343860507011
step: 520, loss: 0.004466413985937834
step: 530, loss: 0.143345907330513

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.


```

    warnings.warn("An input tensor was not cuda.")
epoch 3: dev_f1=0.9335776454420522, f1=0.9333333333333333,
best_f1=0.9324137931034482
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009486760012805462
step: 10, loss: 0.005998565349727869
step: 20, loss: 0.07069248706102371
step: 30, loss: 0.001498477184213698
step: 40, loss: 0.0011760981287807226
step: 50, loss: 0.054705508053302765
step: 60, loss: 0.010828549973666668
step: 70, loss: 0.0035796533338725567
step: 80, loss: 0.013548375107347965
step: 90, loss: 0.015678243711590767
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 100, loss: 0.34552764892578125
step: 110, loss: 0.008983689360320568
step: 120, loss: 0.0012564407661557198
step: 130, loss: 0.05992547422647476
step: 140, loss: 0.009003461338579655
step: 150, loss: 0.0037747228052467108
step: 160, loss: 0.0012537039583548903
step: 170, loss: 0.04286837577819824
step: 180, loss: 0.0267228651791811
step: 190, loss: 0.06456317007541656
step: 200, loss: 0.00277116522192955
step: 210, loss: 0.021870113909244537
step: 220, loss: 0.0033865419682115316
step: 230, loss: 0.1769888550043106
step: 240, loss: 0.008711534552276134
step: 250, loss: 0.0003473286342341453
step: 260, loss: 0.10778908431529999
step: 270, loss: 0.004859219770878553
step: 280, loss: 0.008141164667904377
step: 290, loss: 0.02942889928817749
step: 300, loss: 0.006976725067943335
step: 310, loss: 0.0020627020858228207
step: 320, loss: 0.028556808829307556
step: 330, loss: 0.06811374425888062
step: 340, loss: 0.028173375874757767
step: 350, loss: 0.0735984817147255
step: 360, loss: 0.011686597019433975
step: 370, loss: 0.05612356588244438
step: 380, loss: 0.003615431720390916
step: 390, loss: 0.023337269201874733
step: 400, loss: 0.010194560512900352

```

```

step: 410, loss: 0.030440816655755043
step: 420, loss: 0.014876620844006538
step: 430, loss: 0.016570014879107475
step: 440, loss: 0.048954304307699203
step: 450, loss: 0.045907411724328995
step: 460, loss: 0.0005652441759593785
step: 470, loss: 0.018669484183192253
step: 480, loss: 0.003136839484795928
step: 490, loss: 0.0950993075966835
step: 500, loss: 0.01211352739483118
step: 510, loss: 0.039203185588121414
step: 520, loss: 0.0019587150309234858
step: 530, loss: 0.0009770728647708893
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 4: dev_f1=0.9378794955628212, f1=0.927536231884058,
best_f1=0.927536231884058
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004623860120773315
step: 10, loss: 0.0021487728226929903
step: 20, loss: 0.0015164504293352365
step: 30, loss: 0.0010125539265573025
step: 40, loss: 0.08453620225191116
step: 50, loss: 0.005928165279328823
step: 60, loss: 0.005022380501031876
step: 70, loss: 0.0017998756375163794
step: 80, loss: 0.007891397923231125
step: 90, loss: 0.015241113491356373
step: 100, loss: 0.0015194346196949482
step: 110, loss: 0.010126637294888496
step: 120, loss: 0.007421447895467281
step: 130, loss: 0.0012938460567966104
step: 140, loss: 0.0007944652461446822
step: 150, loss: 0.002133381087332964
step: 160, loss: 0.04067780822515488
step: 170, loss: 0.00939361285418272
step: 180, loss: 0.00719681428745389
step: 190, loss: 0.0002518528199288994
step: 200, loss: 0.004318553488701582
step: 210, loss: 0.004670294001698494
step: 220, loss: 0.000466637487988919

```

```

step: 230, loss: 0.000860933680087328
step: 240, loss: 0.0161743201315403
step: 250, loss: 0.0009167721145786345
step: 260, loss: 0.054360635578632355
step: 270, loss: 0.0018617685418576002
step: 280, loss: 0.02370763197541237
step: 290, loss: 0.0005179433501325548
step: 300, loss: 0.06924908608198166
step: 310, loss: 0.002765106502920389
step: 320, loss: 0.01216869056224823
step: 330, loss: 0.01463270839303732
step: 340, loss: 0.020917480811476707
step: 350, loss: 0.0023530530743300915
step: 360, loss: 0.012308105826377869
step: 370, loss: 0.010095740668475628
step: 380, loss: 0.046426061540842056
step: 390, loss: 0.013631251640617847
step: 400, loss: 0.02523628994822502
step: 410, loss: 0.004787962883710861
step: 420, loss: 0.008900604210793972
step: 430, loss: 0.009727133437991142
step: 440, loss: 0.007607065606862307
step: 450, loss: 0.008380322717130184
step: 460, loss: 0.002135620452463627
step: 470, loss: 0.05023856833577156
step: 480, loss: 0.04601069912314415
step: 490, loss: 0.001455784891732037
step: 500, loss: 0.04504977911710739
step: 510, loss: 0.0014639574801549315
step: 520, loss: 0.008099451661109924
step: 530, loss: 0.017060743644833565
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 5: dev_f1=0.9369863013698629, f1=0.9397260273972603,
best_f1=0.927536231884058
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0021900625433772802
step: 10, loss: 0.037982333451509476
step: 20, loss: 0.0006220860523171723
step: 30, loss: 0.0027453252114355564
step: 40, loss: 0.001176612451672554
step: 50, loss: 0.0010176163632422686
step: 60, loss: 0.0012470957590267062
step: 70, loss: 0.019566496834158897
step: 80, loss: 0.003005199832841754

```

```
step: 90, loss: 0.0014042506227269769
step: 100, loss: 0.019269881770014763
step: 110, loss: 0.1122698113322258
step: 120, loss: 0.004928804002702236
step: 130, loss: 0.007078586611896753
step: 140, loss: 0.03410899266600609
step: 150, loss: 0.0010896389139816165
step: 160, loss: 0.0007196550723165274
step: 170, loss: 0.00020954347564838827
step: 180, loss: 0.004099064506590366
step: 190, loss: 0.06193108484148979
step: 200, loss: 0.00015568961680401117
step: 210, loss: 0.003297062823548913
step: 220, loss: 0.00024369475431740284
step: 230, loss: 0.023745426908135414
step: 240, loss: 0.0004605152935255319
step: 250, loss: 0.0008994229719974101
step: 260, loss: 0.0036076223477721214
step: 270, loss: 0.0008073646458797157
step: 280, loss: 0.06830085813999176
step: 290, loss: 0.009147914126515388
step: 300, loss: 0.00834627915173769
step: 310, loss: 0.0012867264449596405
step: 320, loss: 0.005963935516774654
step: 330, loss: 0.0028724479489028454
step: 340, loss: 0.013505417853593826
step: 350, loss: 0.13364040851593018
step: 360, loss: 0.0006025201291777194
step: 370, loss: 0.16477149724960327
step: 380, loss: 0.046227626502513885
step: 390, loss: 0.012483485974371433
step: 400, loss: 0.0005157947889529169
step: 410, loss: 0.0001843395148171112
step: 420, loss: 0.006774138659238815
step: 430, loss: 0.00017643241153564304
step: 440, loss: 0.07519104331731796
step: 450, loss: 0.0018662138609215617
step: 460, loss: 0.0007405038923025131
step: 470, loss: 0.0035688739735633135
step: 480, loss: 0.008556482382118702
step: 490, loss: 0.00014009277219884098
step: 500, loss: 0.00628227973356843
step: 510, loss: 0.00013987708371132612
step: 520, loss: 0.001781396334990859
step: 530, loss: 0.0018020103452727199
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
epoch 6: dev_f1=0.9282700421940927, f1=0.9212635549269212,  
best_f1=0.927536231884058  
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:  
An input tensor was not cuda.  
    warnings.warn("An input tensor was not cuda.")  
step: 0, loss: 0.004240980371832848  
step: 10, loss: 0.0049459184519946575  
step: 20, loss: 0.01491665281355381  
step: 30, loss: 0.015917565673589706  
step: 40, loss: 0.09556026011705399  
step: 50, loss: 0.004722099285572767  
step: 60, loss: 0.10748974978923798  
step: 70, loss: 0.011969993822276592  
step: 80, loss: 0.05246280878782272  
step: 90, loss: 0.001291819498874247  
step: 100, loss: 0.0019131219014525414  
step: 110, loss: 0.0019178331131115556  
step: 120, loss: 0.0005279323668219149  
step: 130, loss: 0.0005704037030227482  
step: 140, loss: 0.021701108664274216  
step: 150, loss: 0.03277243301272392  
step: 160, loss: 0.001221354934386909  
step: 170, loss: 0.0026301101315766573  
step: 180, loss: 0.000589293020311743  
step: 190, loss: 0.00020535259682219476  
step: 200, loss: 0.006998965982347727  
step: 210, loss: 0.0018066747579723597  
step: 220, loss: 0.0004890907439403236  
step: 230, loss: 0.011205548420548439  
step: 240, loss: 0.034828927367925644  
step: 250, loss: 0.0030098562128841877  
step: 260, loss: 0.026385970413684845  
step: 270, loss: 0.0012695236364379525  
step: 280, loss: 0.037264470010995865  
step: 290, loss: 0.010562785901129246  
step: 300, loss: 0.003830330679193139  
step: 310, loss: 0.001726357964798808  
step: 320, loss: 0.02810649760067463  
step: 330, loss: 0.00023395594325847924  
step: 340, loss: 0.0019107137341052294  
step: 350, loss: 0.0001111478777602315  
step: 360, loss: 0.001536355004645884  
step: 370, loss: 0.011327585205435753  
step: 380, loss: 0.04429682716727257  
step: 390, loss: 9.967525693355128e-05  
step: 400, loss: 0.005416753236204386  
step: 410, loss: 0.0005087046301923692  
step: 420, loss: 0.0033046738244593143
```

```

step: 430, loss: 7.670123159186915e-05
step: 440, loss: 3.267365173087455e-05
step: 450, loss: 0.001345708966255188
step: 460, loss: 0.001270709908567369
step: 470, loss: 0.033507343381643295
step: 480, loss: 0.0011404359247535467
step: 490, loss: 0.00013259818661026657
step: 500, loss: 3.135005317744799e-05
step: 510, loss: 0.003354865126311779
step: 520, loss: 0.042571648955345154
step: 530, loss: 0.0022645611315965652
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 7: dev_f1=0.9381584974805315, f1=0.932666060054595,
best_f1=0.932666060054595
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001190756680443883
step: 10, loss: 0.040320541709661484
step: 20, loss: 0.04929899796843529
step: 30, loss: 0.01348591223359108
step: 40, loss: 0.00028252816991880536
step: 50, loss: 0.0006739685777574778
step: 60, loss: 6.707011198159307e-05
step: 70, loss: 0.04854636266827583
step: 80, loss: 9.687572310213e-05
step: 90, loss: 0.00011394399189157411
step: 100, loss: 0.009744486771523952
step: 110, loss: 0.0006472804234363139
step: 120, loss: 0.003998351749032736
step: 130, loss: 0.0006801420822739601
step: 140, loss: 0.01953926309943199
step: 150, loss: 0.0031872973777353764
step: 160, loss: 0.005073631647974253
step: 170, loss: 0.03549741208553314
step: 180, loss: 0.001187140354886651
step: 190, loss: 0.0038050541188567877
step: 200, loss: 0.0004955691401846707
step: 210, loss: 0.0021694654133170843
step: 220, loss: 0.0013318425044417381
step: 230, loss: 0.003326529636979103
step: 240, loss: 0.010492989793419838

```

```

step: 250, loss: 0.012256285175681114
step: 260, loss: 0.029687825590372086
step: 270, loss: 3.383594594197348e-05
step: 280, loss: 0.011682407930493355
step: 290, loss: 0.0010489748092368245
step: 300, loss: 0.0012962388573214412
step: 310, loss: 0.04588824883103371
step: 320, loss: 0.00016773435345385224
step: 330, loss: 0.042362701147794724
step: 340, loss: 0.0009625889942981303
step: 350, loss: 0.027740752324461937
step: 360, loss: 0.004002054687589407
step: 370, loss: 0.0004689142224378884
step: 380, loss: 0.08269167691469193
step: 390, loss: 0.0002662281331140548
step: 400, loss: 0.03578701615333557
step: 410, loss: 0.0009743004338815808
step: 420, loss: 9.958745067706332e-05
step: 430, loss: 0.0028273628558963537
step: 440, loss: 0.0037281138356775045
step: 450, loss: 0.0008788073901087046
step: 460, loss: 0.010996861383318901
step: 470, loss: 0.0014820233918726444
step: 480, loss: 0.00020520763064268976
step: 490, loss: 0.0001302532764384523
step: 500, loss: 0.006129730027168989
step: 510, loss: 0.01064017042517662
step: 520, loss: 0.0007687755860388279
step: 530, loss: 0.00022421975154429674
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9369453526389537, f1=0.937471051412691,
best_f1=0.932666060054595
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00025554094463586807
step: 10, loss: 0.0007906073005869985
step: 20, loss: 0.020333167165517807
step: 30, loss: 0.005765366833657026
step: 40, loss: 0.009659848175942898
step: 50, loss: 0.0002051472692983225
step: 60, loss: 0.00023574648366775364
step: 70, loss: 0.11524103581905365
step: 80, loss: 0.0002596599224489182
step: 90, loss: 0.0008934583165682852
step: 100, loss: 0.0003381188726052642

```

```

step: 110, loss: 0.00164444942448288202
step: 120, loss: 0.05295851081609726
step: 130, loss: 0.013616942800581455
step: 140, loss: 0.1361706256866455
step: 150, loss: 0.0005431252066046
step: 160, loss: 0.0007947383564896882
step: 170, loss: 0.0001977248175535351
step: 180, loss: 0.0014479800593107939
step: 190, loss: 0.006068028975278139
step: 200, loss: 0.02656598575413227
step: 210, loss: 0.0002371520531596616
step: 220, loss: 0.01862034574151039
step: 230, loss: 0.0002298317849636078
step: 240, loss: 0.0002492883359082043
step: 250, loss: 0.00432057399302721
step: 260, loss: 0.0029066966380923986
step: 270, loss: 0.0019915818702429533
step: 280, loss: 4.6516994189005345e-05
step: 290, loss: 6.822747673140839e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 300, loss: 0.00032719297450967133
step: 310, loss: 1.9486767996568233e-05
step: 320, loss: 0.0015400488628074527
step: 330, loss: 0.0011442664545029402
step: 340, loss: 7.625660509802401e-05
step: 350, loss: 0.038963016122579575
step: 360, loss: 0.03950870782136917
step: 370, loss: 0.00028533689328469336
step: 380, loss: 0.00016603467520326376
step: 390, loss: 0.0011982815340161324
step: 400, loss: 0.001127000548876822
step: 410, loss: 0.00048551277723163366
step: 420, loss: 0.00223811948671937
step: 430, loss: 0.0002371950977249071
step: 440, loss: 0.00012603263894561678
step: 450, loss: 0.00022680862457491457
step: 460, loss: 9.572356793796644e-05
step: 470, loss: 4.039816849399358e-05
step: 480, loss: 0.00025701578124426305
step: 490, loss: 0.002949521644040942
step: 500, loss: 0.041791923344135284
step: 510, loss: 0.11784762889146805
step: 520, loss: 0.0007189117022790015
step: 530, loss: 0.00025196693604812026
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:

```


UserWarning: Please also save or load the state of the optimizer when saving or loading the scheduler.

```
warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 9: dev_f1=0.9425925925925925, f1=0.933456561922366,
best_f1=0.933456561922366
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.03389184921979904
step: 10, loss: 9.862904698820785e-05
step: 20, loss: 0.0006246886332519352
step: 30, loss: 0.00017074766219593585
step: 40, loss: 0.017274409532546997
step: 50, loss: 0.00019005565263796598
step: 60, loss: 0.005994135048240423
step: 70, loss: 0.005273011047393084
step: 80, loss: 0.008664771914482117
step: 90, loss: 0.0017064007697626948
step: 100, loss: 0.017997603863477707
step: 110, loss: 0.0007058435003273189
step: 120, loss: 0.0009853793308138847
step: 130, loss: 0.00185687025077641
step: 140, loss: 0.0006819096161052585
step: 150, loss: 9.341842087451369e-05
step: 160, loss: 0.00139189837500453
step: 170, loss: 8.974086813395843e-05
step: 180, loss: 0.007373816333711147
step: 190, loss: 0.0008214429253712296
step: 200, loss: 0.0028021521866321564
step: 210, loss: 0.00732969306409359
step: 220, loss: 0.0009104680502787232
step: 230, loss: 0.0015812857309356332
step: 240, loss: 0.0003208649577572942
step: 250, loss: 0.0006371171912178397
step: 260, loss: 0.00559623958542943
step: 270, loss: 0.00015867086767684668
step: 280, loss: 1.2963910194230266e-05
step: 290, loss: 3.844866660074331e-05
step: 300, loss: 0.0003726478316821158
step: 310, loss: 0.00036745943361893296
step: 320, loss: 0.0008099671686068177
step: 330, loss: 0.012516035698354244
step: 340, loss: 4.709951099357568e-05
step: 350, loss: 6.148204556666315e-05
step: 360, loss: 5.222222171141766e-05
step: 370, loss: 0.0005452221375890076
step: 380, loss: 3.9726215618429706e-05
step: 390, loss: 2.6019944925792515e-05
```

```

step: 400, loss: 3.9558508433401585e-05
step: 410, loss: 0.00020295722060836852
step: 420, loss: 0.013738340698182583
step: 430, loss: 5.8903588069370016e-05
step: 440, loss: 3.2201183785218745e-05
step: 450, loss: 0.002187171718105674
step: 460, loss: 0.000302973116049543
step: 470, loss: 5.7396155170863494e-05
step: 480, loss: 0.003314525820314884
step: 490, loss: 0.03943716362118721
step: 500, loss: 0.028430089354515076
step: 510, loss: 0.0007926916005089879
step: 520, loss: 0.0002195120177930221
step: 530, loss: 0.00116183131467551
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9424964936886395, f1=0.9353932584269663,
best_f1=0.933456561922366
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005155694088898599
step: 10, loss: 0.023549439385533333
step: 20, loss: 7.290553185157478e-05
step: 30, loss: 0.0008357297047041357
step: 40, loss: 0.0008010892197489738
step: 50, loss: 0.0004083484527654946
step: 60, loss: 0.0003948233206756413
step: 70, loss: 0.000138550836709328
step: 80, loss: 3.419159838813357e-05
step: 90, loss: 0.0011714862193912268
step: 100, loss: 0.00012572079140227288
step: 110, loss: 0.002318182960152626
step: 120, loss: 0.008361061103641987
step: 130, loss: 8.372864976990968e-05
step: 140, loss: 0.00011093377543147653
step: 150, loss: 9.087864600587636e-05
step: 160, loss: 0.004234260879456997
step: 170, loss: 0.046413399279117584
step: 180, loss: 0.000119919961434789
step: 190, loss: 0.0004572541220113635
step: 200, loss: 0.00119551841635257
step: 210, loss: 2.1650865164701827e-05
step: 220, loss: 0.0006429267232306302
step: 230, loss: 0.00591156305745244
step: 240, loss: 4.695840470958501e-05
step: 250, loss: 0.0001140010281233117

```

```

step: 260, loss: 2.7438341930974275e-05
step: 270, loss: 0.0010400473838672042
step: 280, loss: 0.0003126828232780099
step: 290, loss: 0.07151336967945099
step: 300, loss: 0.28942975401878357
step: 310, loss: 0.010487479157745838
step: 320, loss: 0.018028058111667633
step: 330, loss: 0.001759229926392436
step: 340, loss: 0.0030493836384266615
step: 350, loss: 0.026308786123991013
step: 360, loss: 0.0008049693424254656
step: 370, loss: 0.002015984384343028
step: 380, loss: 0.00013263130676932633
step: 390, loss: 0.004443833138793707
step: 400, loss: 0.00015228535630740225
step: 410, loss: 0.00027052167570218444
step: 420, loss: 0.003866345388814807
step: 430, loss: 0.002098826691508293
step: 440, loss: 7.87314202170819e-05
step: 450, loss: 0.0002939805272035301
step: 460, loss: 0.0065780989825725555
step: 470, loss: 0.0012424507876858115
step: 480, loss: 0.0012919335858896375
step: 490, loss: 0.06450755149126053
step: 500, loss: 0.0008582944865338504
step: 510, loss: 0.0038005427923053503
step: 520, loss: 3.6031506169820204e-05
step: 530, loss: 0.0001617572852410376
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 11: dev_f1=0.9428704133766836, f1=0.9333333333333333,
best_f1=0.9333333333333333
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008396698394790292
step: 10, loss: 0.00017487937293481082
step: 20, loss: 5.5942567996680737e-05
step: 30, loss: 0.002103711711242795
step: 40, loss: 0.00015464589523617178
step: 50, loss: 0.009446117095649242
step: 60, loss: 0.0003254749462939799
step: 70, loss: 0.00040580242057330906

```

```
step: 80, loss: 0.0012100253952667117
step: 90, loss: 0.00031236439826898277
step: 100, loss: 0.002017415827140212
step: 110, loss: 0.00013538530038204044
step: 120, loss: 0.001973655540496111
step: 130, loss: 0.0018299202201887965
step: 140, loss: 0.00014032775652594864
step: 150, loss: 0.003493271768093109
step: 160, loss: 0.0017718160524964333
step: 170, loss: 0.003430616809055209
step: 180, loss: 0.0008069159230217338
step: 190, loss: 0.00015488934877794236
step: 200, loss: 6.286401185207069e-05
step: 210, loss: 0.00023288166266866028
step: 220, loss: 2.419739575998392e-05
step: 230, loss: 0.0006025393959134817
step: 240, loss: 0.022127632051706314
step: 250, loss: 0.0001949704746948555
step: 260, loss: 0.000269415439106524
step: 270, loss: 0.09463303536176682
step: 280, loss: 3.060715243918821e-05
step: 290, loss: 0.07414721697568893
step: 300, loss: 0.166495218873024
step: 310, loss: 0.0002790613507386297
step: 320, loss: 0.03980245068669319
step: 330, loss: 2.0737881641252898e-05
step: 340, loss: 0.006073033902794123
step: 350, loss: 0.006060279905796051
step: 360, loss: 0.0014773680595681071
step: 370, loss: 0.00017493720224592835
step: 380, loss: 0.0014386854600161314
step: 390, loss: 0.000500723603181541
step: 400, loss: 0.00033942973823286593
step: 410, loss: 0.007307704072445631
step: 420, loss: 0.013563863933086395
step: 430, loss: 0.0008434297633357346
step: 440, loss: 0.0004968154244124889
step: 450, loss: 0.001749423798173666
step: 460, loss: 0.0005468497984111309
step: 470, loss: 0.008536010049283504
step: 480, loss: 0.00015228141273837537
step: 490, loss: 0.002436043694615364
step: 500, loss: 0.0018393840873613954
step: 510, loss: 0.00014319954789243639
step: 520, loss: 0.12302814424037933
step: 530, loss: 0.00068891845876351
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```

warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 12: dev_f1=0.9437470943747095, f1=0.9398148148148149,
best_f1=0.9398148148148149
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00015794574574101716
step: 10, loss: 0.00015890214126557112
step: 20, loss: 0.018270833417773247
step: 30, loss: 0.0004214822256471962
step: 40, loss: 0.0008573101367801428
step: 50, loss: 0.000221697919187136
step: 60, loss: 0.0006881066947244108
step: 70, loss: 8.136559335980564e-05
step: 80, loss: 2.381082776992116e-05
step: 90, loss: 0.0011566063622012734
step: 100, loss: 0.0006996243610046804
step: 110, loss: 0.0003922336909454316
step: 120, loss: 3.341274714330211e-05
step: 130, loss: 0.002228078432381153
step: 140, loss: 0.0012459984282031655
step: 150, loss: 0.0001255475654033944
step: 160, loss: 8.978893310995772e-05
step: 170, loss: 0.0037586174439638853
step: 180, loss: 4.101052400073968e-05
step: 190, loss: 0.0005487193120643497
step: 200, loss: 0.0018087475327774882
step: 210, loss: 0.002012834884226322
step: 220, loss: 3.055367051274516e-05
step: 230, loss: 0.0019228348974138498
step: 240, loss: 2.4424656658084132e-05
step: 250, loss: 0.026495948433876038
step: 260, loss: 0.00047117078793235123
step: 270, loss: 0.002539969515055418
step: 280, loss: 0.00025873296544887125
step: 290, loss: 0.00010901878704316914
step: 300, loss: 0.0005563012091442943
step: 310, loss: 0.00111066410318017
step: 320, loss: 3.088724770350382e-05
step: 330, loss: 0.009393339045345783
step: 340, loss: 5.7079862017417327e-05
step: 350, loss: 0.014065368101000786
step: 360, loss: 0.000144788486068137
step: 370, loss: 0.0003103697090409696

```

```

step: 380, loss: 0.0004886534879915416
step: 390, loss: 0.003282628720626235
step: 400, loss: 7.723989256191999e-05
step: 410, loss: 5.544837040361017e-05
step: 420, loss: 5.358803173294291e-05
step: 430, loss: 0.00021701610239688307
step: 440, loss: 0.0005635303095914423
step: 450, loss: 0.010293944738805294
step: 460, loss: 8.752997382543981e-05
step: 470, loss: 0.001132402801886201
step: 480, loss: 0.0006689346628263593
step: 490, loss: 0.000111961635411717
step: 500, loss: 0.0003306538856122643
step: 510, loss: 3.0094830435700715e-05
step: 520, loss: 3.371713319211267e-05
step: 530, loss: 3.8840615161461756e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 13: dev_f1=0.9462465245597776, f1=0.9377593360995851,
best_f1=0.9377593360995851
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00020171274081803858
step: 10, loss: 0.00015384118887595832
step: 20, loss: 1.2572741070471238e-05
step: 30, loss: 0.0002747590187937021
step: 40, loss: 0.0001028802216751501
step: 50, loss: 0.0003635766333900392
step: 60, loss: 2.0324539946159348e-05
step: 70, loss: 9.113045234698802e-05
step: 80, loss: 0.00031417427817359567
step: 90, loss: 1.899860944831744e-05
step: 100, loss: 1.5202646864054259e-05
step: 110, loss: 4.264057497493923e-05
step: 120, loss: 4.9076512368628755e-05
step: 130, loss: 0.00018994859419763088
step: 140, loss: 0.0010768716456368566
step: 150, loss: 0.03134359046816826
step: 160, loss: 0.0014710825635120273
step: 170, loss: 6.643379310844466e-05
step: 180, loss: 0.00034856630372814834
step: 190, loss: 0.00013089329877402633

```

```

step: 200, loss: 0.009547458961606026
step: 210, loss: 7.978622306836769e-05
step: 220, loss: 0.11292509734630585
step: 230, loss: 0.001171286217868328
step: 240, loss: 0.0006358759710565209
step: 250, loss: 0.000519138528034091
step: 260, loss: 5.571630390477367e-05
step: 270, loss: 0.0038809904363006353
step: 280, loss: 0.0001586754951858893
step: 290, loss: 0.009790572337806225
step: 300, loss: 0.0006525989738292992
step: 310, loss: 0.0002794662141241133
step: 320, loss: 0.001137174665927887
step: 330, loss: 7.365191413555294e-05
step: 340, loss: 0.0006292558391578496
step: 350, loss: 0.0030990673694759607
step: 360, loss: 0.0016849434468895197
step: 370, loss: 0.019894758239388466
step: 380, loss: 4.460144555196166e-05
step: 390, loss: 0.0002022548287641257
step: 400, loss: 2.029456118179951e-05
step: 410, loss: 3.0667779356008396e-05
step: 420, loss: 0.0001250246714334935
step: 430, loss: 0.0014849897706881166
step: 440, loss: 9.573946044838522e-06
step: 450, loss: 0.000500663067214191
step: 460, loss: 0.030380792915821075
step: 470, loss: 6.181438220664859e-05
step: 480, loss: 0.00015044189058244228
step: 490, loss: 0.0006601723725907505
step: 500, loss: 0.00015918471035547554
step: 510, loss: 0.031484801322221756
step: 520, loss: 2.6536155928624794e-05
step: 530, loss: 0.00047779749729670584
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.943466172381835, f1=0.9390018484288354,
best_f1=0.9377593360995851
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0017492299666628242
step: 10, loss: 0.0006131540285423398
step: 20, loss: 0.0001446038659196347
step: 30, loss: 0.0009593277936801314
step: 40, loss: 2.965173734992277e-05
step: 50, loss: 0.0005509610055014491

```

step: 60, loss: 1.4457684301305562e-05
step: 70, loss: 0.0009596134768798947
step: 80, loss: 4.204772994853556e-05
step: 90, loss: 1.6904903532122262e-05
step: 100, loss: 0.0009320031385868788
step: 110, loss: 0.009346646256744862
step: 120, loss: 0.0015574395656585693
step: 130, loss: 0.00012707301357295364
step: 140, loss: 2.4078697606455535e-05
step: 150, loss: 0.000307718786643818
step: 160, loss: 6.324341666186228e-05
step: 170, loss: 0.00013916839088778943
step: 180, loss: 0.00010423958883620799
step: 190, loss: 0.00026824811357073486
step: 200, loss: 0.0017025923589244485
step: 210, loss: 0.00020049164595548064
step: 220, loss: 0.0001021996940835379
step: 230, loss: 2.7192949346499518e-05
step: 240, loss: 0.00010145718260901049
step: 250, loss: 0.00017955314251594245
step: 260, loss: 0.0031117687467485666
step: 270, loss: 0.004457585047930479
step: 280, loss: 0.0005449176533147693
step: 290, loss: 0.000451851257821545
step: 300, loss: 0.02485058642923832
step: 310, loss: 0.0009132737759500742
step: 320, loss: 8.487395098200068e-05
step: 330, loss: 0.0002506001910660416
step: 340, loss: 0.0002731264685280621
step: 350, loss: 0.0017179478891193867
step: 360, loss: 0.00019055255688726902
step: 370, loss: 1.2885621799796354e-05
step: 380, loss: 0.0002107958571286872
step: 390, loss: 0.00032628330518491566
step: 400, loss: 0.00019762937154155225
step: 410, loss: 0.00012709978909697384
step: 420, loss: 0.0016989518189802766
step: 430, loss: 0.00012937214341945946
step: 440, loss: 0.00012436760880518705
step: 450, loss: 0.0010571781313046813
step: 460, loss: 3.815132367890328e-05
step: 470, loss: 0.00028998011839576066
step: 480, loss: 0.19146020710468292
step: 490, loss: 8.93728356459178e-05
step: 500, loss: 0.001778614241629839
step: 510, loss: 3.7006931961514056e-05
step: 520, loss: 8.90502124093473e-05
step: 530, loss: 1.4740685401193332e-05


```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.944932901434521, f1=0.9376443418013857,
best_f1=0.9377593360995851
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
5742it [00:14, 395.54it/s]
load_f1 = 0.9413970932958275
real_f1 = 0.9400187441424556
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:11, 397.93it/s]

```