ADirty_90_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/Snippext_public
%cd Snippext_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 | 16.53 MiB/s, done.
Resolving deltas: 100% (142/142), 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 1.4 MB/s
Collecting numpy==1.19.2
  Downloading numpy-1.19.2-cp37-cp37m-manylinux2010_x86_64.whl (14.5 MB)
     |                         | 14.5 MB 59.8 MB/s
Collecting regex==2019.12.20
  Downloading regex-2019.12.20-cp37-cp37m-manylinux2010_x86_64.whl (689 kB)
     |                         | 689 kB 53.4 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 58.1 MB/s
Requirement already satisfied: sklearn==0.0 in
/usr/local/lib/python3.7/dist-packages (from -r requirements.txt (line 6)) (0.0)
```

1

```
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 56.1 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 | 12.44 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 27.0 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 74.4 MB/s
Collecting torch==1.4.0
  Downloading torch-1.4.0-cp37-cp37m-manylinux1_x86_64.whl (753.4 MB)
     |                    | 753.4 MB 7.1 kB/s
Collecting tqdm==4.41.0
  Downloading tqdm-4.41.0-py2.py3-none-any.whl (56 kB)
     |                    | 56 kB 4.8 MB/s
Collecting transformers==3.1.0
  Downloading transformers-3.1.0-py3-none-any.whl (884 kB)
     |                    | 884 kB 66.3 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 55.0 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: 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)
Collecting srsly<1.1.0,>=0.1.0
  Downloading srsly-1.0.5-cp37-cp37m-manylinux2014_x86_64.whl (184 kB)
     |                    | 184 kB 70.1 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 63.4 MB/s
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)
Collecting plac<1.2.0,>=0.9.6
  Downloading plac-1.1.3-py2.py3-none-any.whl (20 kB)
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 catalogue<1.1.0,>=0.0.7
  Downloading catalogue-1.0.0-py2.py3-none-any.whl (7.7 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: 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 thinc<7.4.0,>=7.3.0
  Downloading thinc-7.3.1-cp37-cp37m-manylinux1_x86_64.whl (2.2 MB)
     |                    | 2.2 MB 68.3 MB/s
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: 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 sacremoses
  Downloading sacremoses-0.0.53.tar.gz (880 kB)
     |                        | 880 kB 56.4 MB/s
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
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: 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: 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: 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: 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: 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: 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: 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=1449919
sha256=7d6af4980a617702d2e1f7ab4fab0c30a20291ad6cfa8d68c978b41ad0689e58
  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=462aa0d62706de110bf780e824956b6f49d1138c652048037bc97b118e6afa87
  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% (139/139), done.
remote: Total 10085 (delta 98), reused 130 (delta 58), pack-reused 9884
Receiving objects: 100% (10085/10085), 14.95 MiB | 12.84 MiB/s, done.
Resolving deltas: 100% (6903/6903), 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 writeable
Created temporary directory: /tmp/pip-ephem-wheel-cache-mj7ia00_
Created temporary directory: /tmp/pip-req-tracker-mbt492c3
Initialized build tracking at /tmp/pip-req-tracker-mbt492c3
Created build tracker: /tmp/pip-req-tracker-mbt492c3
Entered build tracker: /tmp/pip-req-tracker-mbt492c3
```

```
Created temporary directory: /tmp/pip-install-qnxfsksp
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-pqj5iuau
  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-
mbt492c3'
    Running setup.py (path:/tmp/pip-req-build-pqj5iuau/setup.py) egg_info for
package from file:///content/ditto/apex
    Created temporary directory: /tmp/pip-pip-egg-info-y0hmfa86
    Running command python setup.py egg_info



    torch.__version__  = 1.4.0



    running egg_info
    creating /tmp/pip-pip-egg-info-y0hmfa86/apex.egg-info
    writing /tmp/pip-pip-egg-info-y0hmfa86/apex.egg-info/PKG-INFO
    writing dependency_links to /tmp/pip-pip-egg-info-y0hmfa86/apex.egg-
info/dependency_links.txt
    writing top-level names to /tmp/pip-pip-egg-info-y0hmfa86/apex.egg-
info/top_level.txt
    writing manifest file '/tmp/pip-pip-egg-info-y0hmfa86/apex.egg-
info/SOURCES.txt'
    adding license file 'LICENSE'
    writing manifest file '/tmp/pip-pip-egg-info-y0hmfa86/apex.egg-
info/SOURCES.txt'
  Source in /tmp/pip-req-build-pqj5iuau 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-mbt492c3'
Created temporary directory: /tmp/pip-unpack-f4ykhv76
Building wheels for collected packages: apex
  Created temporary directory: /tmp/pip-wheel-xgi3ju9a
  Building wheel for apex (setup.py) …   Destination directory:
/tmp/pip-wheel-xgi3ju9a
  Running command /usr/bin/python3 -u -c 'import io, os, sys, setuptools,
tokenize; sys.argv[0] = '"'"'/tmp/pip-req-build-pqj5iuau/setup.py'"'"';
```

```
__file__='"'"'/tmp/pip-req-build-pqj5iuau/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-xgi3ju9a


  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/_batchsampler.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_pipelining_with_i
nterleaving.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_pipelining.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_pipelining_withou
t_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/_batchsampler.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_pipelin
ing_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_pipelining.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_pipelin
ing_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-xgi3ju9a/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/_batchsampler.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=9bfca4d2e42f1c2876d05605a1349559e0ee5846484543ef3949f30409e9731f
  Stored in directory: /tmp/pip-ephem-wheel-cache-
mj7ia00_/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-mbt492c3'
/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 31.5 MB/s
Collecting awscli
  Downloading awscli-1.25.46-py3-none-any.whl (3.9 MB)
     | | 3.9 MB 60.1 MB/s
Collecting s3transfer<0.7.0,>=0.6.0
  Downloading s3transfer-0.6.0-py3-none-any.whl (79 kB)
     | | 79 kB 6.9 MB/s
Collecting colorama<0.4.5,>=0.2.5
  Downloading colorama-0.4.4-py2.py3-none-any.whl (16 kB)
Collecting rsa<4.8,>=3.1.2
  Downloading rsa-4.7.2-py3-none-any.whl (34 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 68.0 MB/s
Collecting botocore==1.27.46
  Downloading botocore-1.27.46-py3-none-any.whl (9.0 MB)
     | | 9.0 MB 61.6 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: numpy in /usr/local/lib/python3.7/dist-
packages (from torch==1.7.1) (1.19.2)
Requirement already satisfied: typing-extensions in
/usr/local/lib/python3.7/dist-packages (from torch==1.7.1) (4.1.1)
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 | 17.87 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/ADirty_90_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, 395kB/s]
Downloading: 100% 232k/232k [00:00<00:00, 348kB/s]
Downloading: 100% 268M/268M [00:03<00:00, 69.8MB/s]
step: 0, loss: 0.8758458495140076
epoch 1: dev_f1=0.3146067415730337, f1=0.2947368421052632,
best_f1=0.2947368421052632
step: 0, loss: 0.36972710490226746
epoch 2: dev_f1=0.2978723404255319, f1=0.2828282828282828,
best_f1=0.2947368421052632
step: 0, loss: 0.3467859625816345
epoch 3: dev_f1=0.35294117647058826, f1=0.34374999999999994,
best_f1=0.34374999999999994
```

```
step: 0, loss: 0.358087420463562
epoch 4: dev_f1=0.358974358974359, f1=0.3170731707317073,
best_f1=0.3170731707317073
step: 0, loss: 0.259922593832016
epoch 5: dev_f1=0.43076923076923085, f1=0.32876712328767127,
best_f1=0.32876712328767127
step: 0, loss: 0.25525233149528503
epoch 6: dev_f1=0.45161290322580644, f1=0.38095238095238093,
best_f1=0.38095238095238093
step: 0, loss: 0.26379281282424927
epoch 7: dev_f1=0.456140350877193, f1=0.37931034482758624,
best_f1=0.37931034482758624
step: 0, loss: 0.3414508104324341
epoch 8: dev_f1=0.456140350877193, f1=0.38596491228070184,
best_f1=0.37931034482758624
step: 0, loss: 0.2124498039484024
epoch 9: dev_f1=0.4727272727272727, f1=0.38596491228070184,
best_f1=0.38596491228070184
step: 0, loss: 0.24804510176181793
epoch 10: dev_f1=0.55, f1=0.3902439024390244, best_f1=0.3902439024390244
step: 0, loss: 0.1899382472038269
epoch 11: dev_f1=0.5714285714285714, f1=0.38095238095238093,
best_f1=0.38095238095238093
step: 0, loss: 0.2318941205739975
epoch 12: dev_f1=0.6, f1=0.3902439024390244, best_f1=0.3902439024390244
step: 0, loss: 0.1608196645975113
epoch 13: dev_f1=0.5714285714285714, f1=0.37209302325581395,
best_f1=0.3902439024390244
step: 0, loss: 0.16963538527488708
epoch 14: dev_f1=0.5853658536585367, f1=0.37209302325581395,
best_f1=0.3902439024390244
step: 0, loss: 0.2338300347328186
epoch 15: dev_f1=0.5853658536585367, f1=0.37209302325581395,
best_f1=0.3902439024390244
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 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: 0.8142889142036438
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.49180868268013
step: 20, loss: 0.5944226384162903
step: 30, loss: 0.4981488585472107
step: 40, loss: 0.41197651624679565
step: 50, loss: 0.29177790880203247
step: 60, loss: 0.1782059222459793
step: 70, loss: 0.15441790223121643
step: 80, loss: 0.18717901408672333
step: 90, loss: 0.013776328414678574
step: 100, loss: 0.10326925665140152
step: 110, loss: 0.035572364926338196
step: 120, loss: 0.07247243821620941
step: 130, loss: 0.018750831484794617
step: 140, loss: 0.12470116466283798
step: 150, loss: 0.10674717277288437
step: 160, loss: 0.1314665675163269
step: 170, loss: 0.005865129642188549
step: 180, loss: 0.04757256433367729
step: 190, loss: 0.09981340169906616
step: 200, loss: 0.05021965503692627
step: 210, loss: 0.01379481516778469
step: 220, loss: 0.007516228593885899
step: 230, loss: 0.004926974419504404
/usr/local/lib/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=0.9610678531701891, f1=0.9552572706935123,
best_f1=0.9552572706935123
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00555738341063261
step: 10, loss: 0.0029913908801972866
step: 20, loss: 0.055199433118104935
step: 30, loss: 0.0027159410528838634
step: 40, loss: 0.038023363798856735
step: 50, loss: 0.0033961590379476547
step: 60, loss: 0.07621600478887558
step: 70, loss: 0.02917417138814926
step: 80, loss: 0.09467590600252151
step: 90, loss: 0.011628801003098488
step: 100, loss: 0.023550035431981087
step: 110, loss: 0.0682438388466835
step: 120, loss: 0.030039725825190544
step: 130, loss: 0.017003901302814484
step: 140, loss: 0.012667670845985413
step: 150, loss: 0.0049148923717439175
step: 160, loss: 0.008284605108201504
step: 170, loss: 0.12295383214950562
step: 180, loss: 0.008933817967772484
step: 190, loss: 0.06980452686548233
step: 200, loss: 0.004022871609777212
step: 210, loss: 0.02787526324391365
step: 220, loss: 0.0011556862154975533
step: 230, loss: 0.12318331003189087
/usr/local/lib/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=0.9713024282560706, f1=0.968236582694414,
best_f1=0.968236582694414
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10003937780857086
```

```
step: 10, loss: 0.007901323027908802
step: 20, loss: 0.022294480353593826
step: 30, loss: 0.016888627782464027
step: 40, loss: 0.048129402101039886
step: 50, loss: 0.006756053306162357
step: 60, loss: 0.0005643513868562877
step: 70, loss: 0.0006783962016925216
step: 80, loss: 0.0018511834787204862
step: 90, loss: 0.037195466458797455
step: 100, loss: 0.0027715894393622875
step: 110, loss: 0.03426079824566841
step: 120, loss: 0.0026426343247294426
step: 130, loss: 0.0008360648062080145
step: 140, loss: 0.0032264194451272488
step: 150, loss: 0.003088914090767503
step: 160, loss: 0.0014462261460721493
step: 170, loss: 0.012853261083364487
step: 180, loss: 0.002427280182018876
step: 190, loss: 0.013124440796673298
step: 200, loss: 0.030656039714813232
step: 210, loss: 0.01658695563673973
step: 220, loss: 0.0030697593465447426
step: 230, loss: 0.1370357871055603
/usr/local/lib/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=0.9729119638826186, f1=0.9672316384180792,
best_f1=0.9672316384180792
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.039917293936014175
step: 10, loss: 0.0006442363373935223
step: 20, loss: 0.0007308863569051027
step: 30, loss: 0.0003705599228851497
step: 40, loss: 0.006207942962646484
step: 50, loss: 0.009516028687357903
step: 60, loss: 0.015055336989462376
step: 70, loss: 0.006693728733807802
step: 80, loss: 0.03277042508125305
step: 90, loss: 0.0034422925673425198
step: 100, loss: 0.0005191023228690028
step: 110, loss: 0.022816846147179604
step: 120, loss: 0.016175301745533943
```

```
step: 130, loss: 0.12608769536018372
step: 140, loss: 0.0007224288419820368
step: 150, loss: 0.0021273281890898943
step: 160, loss: 0.0036488487385213375
step: 170, loss: 0.020747562870383263
step: 180, loss: 0.07145778089761734
step: 190, loss: 0.03493494540452957
step: 200, loss: 0.002295959508046508
step: 210, loss: 0.018321696668863297
step: 220, loss: 0.0006953920819796622
step: 230, loss: 0.0005862021935172379
/usr/local/lib/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=0.9799107142857142, f1=0.9709821428571428,
best_f1=0.9709821428571428
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007125265547074378
step: 10, loss: 0.00038489318103529513
step: 20, loss: 0.10332272946834564
step: 30, loss: 0.00024281696823891252
step: 40, loss: 0.05540768802165985
step: 50, loss: 0.0010963438544422388
step: 60, loss: 0.0013539429055526853
step: 70, loss: 0.00760737806558609
step: 80, loss: 0.0011544703738763928
step: 90, loss: 0.011554170399904251
step: 100, loss: 0.09446156769990921
step: 110, loss: 0.06305515021085739
step: 120, loss: 0.0616597943007946
step: 130, loss: 0.11120518296957016
step: 140, loss: 0.039398010820150375
step: 150, loss: 0.0004869753902312368
step: 160, loss: 0.004968450404703617
step: 170, loss: 0.028004970401525497
step: 180, loss: 0.0013614615891128778
step: 190, loss: 0.0025779958814382553
step: 200, loss: 0.01900029554963112
step: 210, loss: 0.0004999699885956943
step: 220, loss: 0.00026435701875016093
step: 230, loss: 0.00030603414052166045
/usr/local/lib/python3.7/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=0.9733333333333333, f1=0.9656699889258028,
best_f1=0.9709821428571428
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007906832615844905
step: 10, loss: 0.00028997010667808354
step: 20, loss: 0.007667347323149443
step: 30, loss: 0.00037814694223922547
step: 40, loss: 0.0007448080577887595
step: 50, loss: 0.0022967832628637552
step: 60, loss: 0.01802666485309601
step: 70, loss: 0.020016318187117577
step: 80, loss: 0.00023033114848658442
step: 90, loss: 0.0004533221072051674
step: 100, loss: 0.00047745046322233975
step: 110, loss: 0.027483507990837097
step: 120, loss: 0.00035151338670402765
step: 130, loss: 0.0004898300394415855
step: 140, loss: 0.009537864476442337
step: 150, loss: 0.00017082845442928374
step: 160, loss: 0.0011775599559769034
step: 170, loss: 0.00013232413039077073
step: 180, loss: 0.00022349249047692865
step: 190, loss: 0.0006646704860031605
step: 200, loss: 0.0007502930238842964
step: 210, loss: 0.00017179589485749602
step: 220, loss: 0.00028361883596517146
step: 230, loss: 0.00015690621512476355
/usr/local/lib/python3.7/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=0.971815107102593, f1=0.9695603156708005,
best_f1=0.9709821428571428
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02997007593512535
step: 10, loss: 0.0004734852409455925
step: 20, loss: 0.0014126747846603394
step: 30, loss: 0.00022671103943139315
step: 40, loss: 0.01552001852542162
step: 50, loss: 0.00011145744065288454
step: 60, loss: 0.04879016429185867
step: 70, loss: 0.0004865913069806993
step: 80, loss: 0.00950525514781475
```

```
step: 90, loss: 0.00032351480331271887
step: 100, loss: 0.08838838338851929
step: 110, loss: 0.001556105213239789
step: 120, loss: 0.03574354946613312
step: 130, loss: 0.021453583613038063
step: 140, loss: 0.00030503360903821886
step: 150, loss: 0.00019257277017459273
step: 160, loss: 0.00017914340423885733
step: 170, loss: 0.00048041949048638344
step: 180, loss: 0.00010686242603696883
step: 190, loss: 0.0001629893231438473
step: 200, loss: 0.0001224736770382151
step: 210, loss: 8.842880197335035e-05
step: 220, loss: 0.0003298823721706867
step: 230, loss: 0.0003951615944970399
/usr/local/lib/python3.7/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=0.9752252252252253, f1=0.9753363228699552,
best_f1=0.9709821428571428
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0054892729967832565
step: 10, loss: 0.014757181517779827
step: 20, loss: 0.00015064277977721699
step: 30, loss: 0.00012217488256283104
step: 40, loss: 0.00048663149937056005
step: 50, loss: 0.00011830509902210906
step: 60, loss: 0.0034436634741723537
step: 70, loss: 0.0006301270914264023
step: 80, loss: 0.005589242093265057
step: 90, loss: 0.0001714328973321244
step: 100, loss: 0.00017655998817645013
step: 110, loss: 0.00041175153455697
step: 120, loss: 0.00041864049853757024
step: 130, loss: 0.014788421802222729
step: 140, loss: 0.009452307596802711
step: 150, loss: 0.0017736005829647183
step: 160, loss: 0.001851081964559853
step: 170, loss: 0.0057171983644366264
step: 180, loss: 0.001662139780819416
step: 190, loss: 0.00023732132103759795
step: 200, loss: 0.0008566759061068296
step: 210, loss: 0.00020195619435980916
step: 220, loss: 0.0010920015629380941
step: 230, loss: 0.01146464329212904
/usr/local/lib/python3.7/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=0.9788182831661093, f1=0.970917225950783,
best_f1=0.9709821428571428
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00039909259066917
step: 10, loss: 0.03134036436676979
step: 20, loss: 0.0014953904319554567
step: 30, loss: 0.00035049268626607955
step: 40, loss: 0.0002249896206194535
step: 50, loss: 0.0003749468014575541
step: 60, loss: 0.00013213434431236237
step: 70, loss: 0.0006747626466676593
step: 80, loss: 0.009571300819516182
step: 90, loss: 0.002147089224308729
step: 100, loss: 0.00013197505904827267
step: 110, loss: 0.0005290087428875268
step: 120, loss: 0.00039730226853862405
step: 130, loss: 0.00021011418721172959
step: 140, loss: 0.0008427641005255282
step: 150, loss: 6.0666010540444404e-05
step: 160, loss: 0.00031189803848974407
step: 170, loss: 9.970007522497326e-05
step: 180, loss: 0.0001862876961240545
step: 190, loss: 6.054605182725936e-05
step: 200, loss: 0.0029794303700327873
step: 210, loss: 0.0011961165582761168
step: 220, loss: 0.03131268918514252
step: 230, loss: 0.008282411843538284
/usr/local/lib/python3.7/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=0.9727891156462585, f1=0.9708520179372198,
best_f1=0.9709821428571428
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03950995206832886
step: 10, loss: 6.949830276425928e-05
step: 20, loss: 9.401289571542293e-05
step: 30, loss: 0.0010000684997066855
step: 40, loss: 5.4400559747591615e-05
step: 50, loss: 0.00027439510449767113
step: 60, loss: 0.0004892419092357159
step: 70, loss: 9.832423529587686e-05
step: 80, loss: 0.00014507408195640892
```

```
step: 90, loss: 8.630954835098237e-05
step: 100, loss: 9.254073665942997e-05
step: 110, loss: 0.00012388468894641846
step: 120, loss: 0.08294837921857834
step: 130, loss: 0.00011916028597624972
step: 140, loss: 0.006121920421719551
step: 150, loss: 8.907353185350075e-05
step: 160, loss: 0.005939075723290443
step: 170, loss: 0.00029750948306173086
step: 180, loss: 0.0015653319424018264
step: 190, loss: 0.00031322753056883881
step: 200, loss: 8.832794992486015e-05
step: 210, loss: 0.0001327931968262419
step: 220, loss: 0.0009102686308324337
step: 230, loss: 0.00017794428276829422
/usr/local/lib/python3.7/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=0.9740112994350283, f1=0.968609865470852,
best_f1=0.9709821428571428
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00058766856091097
step: 10, loss: 7.160293171182275e-05
step: 20, loss: 5.9995461924700066e-05
step: 30, loss: 0.0003603781806305051
step: 40, loss: 0.0025560881476849318
step: 50, loss: 0.003230989445000887
step: 60, loss: 0.00022286377497948706
step: 70, loss: 0.0012856151442974806
step: 80, loss: 0.005226234905421734
step: 90, loss: 9.269416477764025e-05
step: 100, loss: 0.00020424378453753889
step: 110, loss: 0.00032692719832994044
step: 120, loss: 0.00016894318105187267
step: 130, loss: 7.632874621776864e-05
step: 140, loss: 0.0019005637150257826
step: 150, loss: 5.9674021031241864e-05
step: 160, loss: 8.846252603689209e-05
step: 170, loss: 0.015494785271584988
step: 180, loss: 0.00029996619559824467
step: 190, loss: 0.00012970523675903678
step: 200, loss: 0.0001440651540178805685
step: 210, loss: 7.183915295172483e-05
step: 220, loss: 0.0004304400354158133
step: 230, loss: 0.0077626388520002365
/usr/local/lib/python3.7/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=0.976324689966178, f1=0.9696287964004499,
best_f1=0.9709821428571428
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00040585504029877484
step: 10, loss: 0.0007573022157885134
step: 20, loss: 0.00011540721607161686
step: 30, loss: 0.0010408023372292519
step: 40, loss: 0.00015952176181599498
step: 50, loss: 0.00043933052802458405
step: 60, loss: 0.001164728426374495
step: 70, loss: 0.00027078803395852447
step: 80, loss: 5.636879723169841e-05
step: 90, loss: 4.864498259848915e-05
step: 100, loss: 5.245423017186113e-05
step: 110, loss: 8.680626342538744e-05
step: 120, loss: 0.00013195117935538292
step: 130, loss: 8.448510925518349e-05
step: 140, loss: 4.841848567593843e-05
step: 150, loss: 6.318598752841353e-05
step: 160, loss: 0.0002523886214476079
step: 170, loss: 0.00012506931670941412
step: 180, loss: 0.0001418698375346139
step: 190, loss: 6.34747848380357e-05
step: 200, loss: 0.00042202157783322036
step: 210, loss: 0.0016459141625091434
step: 220, loss: 0.00026536730001680553
step: 230, loss: 0.0001940123038366437
/usr/local/lib/python3.7/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=0.9751131221719457, f1=0.9685393258426966,
best_f1=0.9709821428571428
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0011966582387685776
step: 10, loss: 0.00017168701742775738
step: 20, loss: 9.026558109326288e-05
step: 30, loss: 0.002538295928388834
step: 40, loss: 0.008202294819056988
step: 50, loss: 7.334292604355142e-05
step: 60, loss: 0.00041673253872431815
step: 70, loss: 6.710334128001705e-05
step: 80, loss: 5.88548973610159e-05
```

```
step: 90, loss: 4.722612356999889e-05
step: 100, loss: 0.00011495008948259056
step: 110, loss: 9.696668712422252e-05
step: 120, loss: 0.0002285892842337489
step: 130, loss: 7.618977542733774e-05
step: 140, loss: 7.743102469248697e-05
step: 150, loss: 0.0009457563864998519
step: 160, loss: 5.6175762438215315e-05
step: 170, loss: 0.00012348213931545615
step: 180, loss: 0.00013321386359166354
step: 190, loss: 5.534174124477431e-05
step: 200, loss: 4.223092400934547e-05
step: 210, loss: 7.505629037041217e-05
step: 220, loss: 0.00013852967822458595
step: 230, loss: 5.228376539889723e-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=0.9773242630385486, f1=0.9751131221719457,
best_f1=0.9709821428571428
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00018999945314135402
step: 10, loss: 4.830619946005754e-05
step: 20, loss: 0.0007540213991887867
step: 30, loss: 0.0002239453897345811
step: 40, loss: 3.789096081163734e-05
step: 50, loss: 0.00022980596986599267
step: 60, loss: 0.00012265128316357732
step: 70, loss: 6.18315243627876e-05
step: 80, loss: 9.989891259465367e-05
step: 90, loss: 0.00017867414862848818
step: 100, loss: 0.00413236441090703
step: 110, loss: 0.00012041905574733391
step: 120, loss: 6.616466998821124e-05
step: 130, loss: 9.155623411061242e-05
step: 140, loss: 8.296371379401535e-05
step: 150, loss: 8.006940333871171e-05
step: 160, loss: 2.8627757274080068e-05
step: 170, loss: 5.7847180869430304e-05
step: 180, loss: 8.542197610950097e-05
step: 190, loss: 0.0002279606560477987
step: 200, loss: 0.06457488983869553
step: 210, loss: 6.826782191637903e-05
step: 220, loss: 8.848823927110061e-05
step: 230, loss: 2.211297396570444e-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=0.9762174405436014, f1=0.9728506787330317,
best_f1=0.9709821428571428
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00031843080068938434
step: 10, loss: 0.0002680267207324505
step: 20, loss: 8.867026190273464e-05
step: 30, loss: 6.143259815871716e-05
step: 40, loss: 9.618565673008561e-05
step: 50, loss: 6.843972369097173e-05
step: 60, loss: 7.479610212612897e-05
step: 70, loss: 4.672394788940437e-05
step: 80, loss: 0.00010064426169265062
step: 90, loss: 5.1069637265754864e-05
step: 100, loss: 0.00012344463902991265
step: 110, loss: 7.492150325560942e-05
step: 120, loss: 0.00019155156041961163
step: 130, loss: 8.963829895947129e-05
step: 140, loss: 0.0001723897730698809
step: 150, loss: 0.00010896188905462623
step: 160, loss: 5.4331685532815754e-05
step: 170, loss: 3.326169098727405e-05
step: 180, loss: 8.266133954748511e-05
step: 190, loss: 0.00025107571855187416
step: 200, loss: 6.323319394141436e-05
step: 210, loss: 0.042441561818122864
step: 220, loss: 4.266308678779751e-05
step: 230, loss: 5.8770045143319294e-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=0.9774266365688488, f1=0.9729119638826186,
best_f1=0.9709821428571428
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, 229.39it/s]
load_f1 = 0.9765886287625419
real_f1 = 0.9776286353467561
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, 250.39it/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, 398kB/s]
Downloading: 100% 232k/232k [00:00<00:00, 335kB/s]
Downloading: 100% 268M/268M [00:04<00:00, 53.7MB/s]
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: 0.7972330451011658
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4509055018424988
step: 20, loss: 0.5209246873855591
step: 30, loss: 0.3583526909351349
step: 40, loss: 0.4215586185455322
step: 50, loss: 0.2178347259759903
step: 60, loss: 0.5026189088821411
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 70, loss: 0.29180771112442017
step: 80, loss: 0.1716846525669098
step: 90, loss: 0.21593064069747925
step: 100, loss: 0.265526682138443
step: 110, loss: 0.05951515585184097
step: 120, loss: 0.12155251950025558
step: 130, loss: 0.04002678766846657
step: 140, loss: 0.2862164378166199
step: 150, loss: 0.03858337923884392
step: 160, loss: 0.17584338784217834
step: 170, loss: 0.36656996607780457
step: 180, loss: 0.2145356833934784
step: 190, loss: 0.08515206724405289
step: 200, loss: 0.1082833781838417
step: 210, loss: 0.04875454679131508
step: 220, loss: 0.1137450560927391
step: 230, loss: 0.1963750123977661
step: 240, loss: 0.11365848779678345
step: 250, loss: 0.12245722115039825
step: 260, loss: 0.043314557522535324
step: 270, loss: 0.10093223303556442
step: 280, loss: 0.13918781280517578
step: 290, loss: 0.057606346905231476
step: 300, loss: 0.08241718262434006
step: 310, loss: 0.14894826710224152
step: 320, loss: 0.1522538661956787
step: 330, loss: 0.130212664604187
step: 340, loss: 0.21439777314662933
step: 350, loss: 0.04707837477326393
step: 360, loss: 0.12151370942592621
step: 370, loss: 0.16522778570652008
step: 380, loss: 0.1154554933309555
step: 390, loss: 0.23881910741329193
step: 400, loss: 0.024518698453903198

```
step: 410, loss: 0.157953143119812
step: 420, loss: 0.03470059111714363
step: 430, loss: 0.15410387516021729
step: 440, loss: 0.3413172662258148
step: 450, loss: 0.03450804576277733
step: 460, loss: 0.08836077898740768
step: 470, loss: 0.2619362771511078
step: 480, loss: 0.2443220317363739
step: 490, loss: 0.1972866952419281
step: 500, loss: 0.01764172501862049
step: 510, loss: 0.13037003576755524
step: 520, loss: 0.12284542620182037
step: 530, loss: 0.2106618881225586
/usr/local/lib/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=0.8927078495123083, f1=0.8921658986175115,
best_f1=0.8921658986175115
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.19091400504112244
step: 10, loss: 0.15337303280830383
step: 20, loss: 0.14216731488704681
step: 30, loss: 0.0713086724281311
step: 40, loss: 0.026237810030579567
step: 50, loss: 0.06962107867002487
step: 60, loss: 0.27664002776145935
step: 70, loss: 0.2518520951271057
step: 80, loss: 0.04575943574309349
step: 90, loss: 0.07681728154420853
step: 100, loss: 0.33020976185798645
step: 110, loss: 0.11175709217786789
step: 120, loss: 0.147967129945755
step: 130, loss: 0.23040643334388733
step: 140, loss: 0.06516799330711365
step: 150, loss: 0.024121925234794617
step: 160, loss: 0.15453599393367767
step: 170, loss: 0.06803379207849503
step: 180, loss: 0.0208150465041399
step: 190, loss: 0.06205582618713379
step: 200, loss: 0.11239083856344223
step: 210, loss: 0.09324543178081512
step: 220, loss: 0.1311739683151245
```

```
step: 230, loss: 0.0695868656039238
step: 240, loss: 0.10863907635211945
step: 250, loss: 0.07815735787153244
step: 260, loss: 0.055469244718551636
step: 270, loss: 0.06026845797896385
step: 280, loss: 0.09962661564350128
step: 290, loss: 0.06499917805194855
step: 300, loss: 0.05474173277616501
step: 310, loss: 0.08669169247150421
step: 320, loss: 0.153006449341774
step: 330, loss: 0.06468402594327927
step: 340, loss: 0.0649450421333313
step: 350, loss: 0.07362443953752518
step: 360, loss: 0.2005264014005661
step: 370, loss: 0.05847134068608284
step: 380, loss: 0.07581474632024765
step: 390, loss: 0.0588541105389595
step: 400, loss: 0.11096438765525818
step: 410, loss: 0.022338995710015297
step: 420, loss: 0.0530117005109787
step: 430, loss: 0.06342471390962601
step: 440, loss: 0.01620299555361271
step: 450, loss: 0.1415419727563858
step: 460, loss: 0.2358543425798416
step: 470, loss: 0.09415559470653534
step: 480, loss: 0.27965155243873596
step: 490, loss: 0.11158712208271027
step: 500, loss: 0.044513896107673645
step: 510, loss: 0.1399642527103424
step: 520, loss: 0.05614670366048813
step: 530, loss: 0.2574693262577057
/usr/local/lib/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=0.9075011504832029, f1=0.9024278515803941,
best_f1=0.9024278515803941
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08324822783470154
step: 10, loss: 0.1464596837759018
step: 20, loss: 0.06967541575431824
step: 30, loss: 0.16113217175006866
step: 40, loss: 0.012102157808840275
```

```
step: 50, loss: 0.03615861386060715
step: 60, loss: 0.010997907258570194
step: 70, loss: 0.025018993765115738
step: 80, loss: 0.07728356122970581
step: 90, loss: 0.20802423357963562
step: 100, loss: 0.013189252465963364
step: 110, loss: 0.04238142445683479
step: 120, loss: 0.12466748058795929
step: 130, loss: 0.015552029013633728
step: 140, loss: 0.034241124987602234
step: 150, loss: 0.03155384212732315
step: 160, loss: 0.02117367461323738
step: 170, loss: 0.029078206047415733
step: 180, loss: 0.07371551543474197
step: 190, loss: 0.03505278378725052
step: 200, loss: 0.09377507865428925
step: 210, loss: 0.041702110320329666
step: 220, loss: 0.04981641471385956
step: 230, loss: 0.04625977948307991
step: 240, loss: 0.011381284333765507
step: 250, loss: 0.040541764348745346
step: 260, loss: 0.005174242425709963
step: 270, loss: 0.005971590522676706
step: 280, loss: 0.04196004197001457
step: 290, loss: 0.09920982271432877
step: 300, loss: 0.11706849932670593
step: 310, loss: 0.12434443831443787
step: 320, loss: 0.03466993570327759
step: 330, loss: 0.011146286502480507
step: 340, loss: 0.035296447575092316
step: 350, loss: 0.053615324199199677
step: 360, loss: 0.011292262002825737
step: 370, loss: 0.010468779131770134
step: 380, loss: 0.026960868388414383
step: 390, loss: 0.10463028401136398
step: 400, loss: 0.0675613284111023
step: 410, loss: 0.16702453792095184
step: 420, loss: 0.11409929394721985
step: 430, loss: 0.024519864469766617
step: 440, loss: 0.09460975974798203
step: 450, loss: 0.1778477281332016
step: 460, loss: 0.31365901231765747
step: 470, loss: 0.03518729656934738
step: 480, loss: 0.05305511876940727
step: 490, loss: 0.06137982755899429
step: 500, loss: 0.09720521420240402
step: 510, loss: 0.07938829809427261
step: 520, loss: 0.015086636878550053
```

```
step: 530, loss: 0.00918387621641159
/usr/local/lib/python3.7/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=0.9047619047619048, f1=0.8993476234855545,
best_f1=0.9024278515803941
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0064465999603271484
step: 10, loss: 0.008888421580195427
step: 20, loss: 0.03140440955758095
step: 30, loss: 0.04197661206126213
step: 40, loss: 0.009947292506694794
step: 50, loss: 0.02573617734014988
step: 60, loss: 0.0014169467613101006
step: 70, loss: 0.027948234230279922
step: 80, loss: 0.014770531095564365
step: 90, loss: 0.09384023398160934
step: 100, loss: 0.09134475141763687
step: 110, loss: 0.018896937370300293
step: 120, loss: 0.003013978712260723
step: 130, loss: 0.00706806406378746
step: 140, loss: 0.010748909786343575
step: 150, loss: 0.09958356618881226
step: 160, loss: 0.017171721905469894
step: 170, loss: 0.0471026636660099
step: 180, loss: 0.02335173450410366
step: 190, loss: 0.013233669102191925
step: 200, loss: 0.01628018356859684
step: 210, loss: 0.17691904306411743
step: 220, loss: 0.01093947235494852
step: 230, loss: 0.2855607271194458
step: 240, loss: 0.007668275386095047
step: 250, loss: 0.0044075255282223225
step: 260, loss: 0.10717009752988815
step: 270, loss: 0.1271342635154724
step: 280, loss: 0.004984704311937094
step: 290, loss: 0.012570780701935291
step: 300, loss: 0.027631357312202454
step: 310, loss: 0.011842364445328712
step: 320, loss: 0.07040082663297653
step: 330, loss: 0.12961870431900024
step: 340, loss: 0.08488541841506958
step: 350, loss: 0.008465554565191269
step: 360, loss: 0.05080005154013634
step: 370, loss: 0.0402839258313179
step: 380, loss: 0.0081876767779925823
```

```
step: 390, loss: 0.047734469175338745
step: 400, loss: 0.04015196114778519
step: 410, loss: 0.1401553452014923
step: 420, loss: 0.03009692393243313
step: 430, loss: 0.01179924700409174
step: 440, loss: 0.10535946488380432
step: 450, loss: 0.08533062040805817
step: 460, loss: 0.001170016941614449
step: 470, loss: 0.012154452502727509
step: 480, loss: 0.01139246765524149
step: 490, loss: 0.0634010061621666
step: 500, loss: 0.014160208404064178
step: 510, loss: 0.12132467329502106
step: 520, loss: 0.1336139738559723
step: 530, loss: 0.004964292980730534
/usr/local/lib/python3.7/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=0.90625, f1=0.9032562529495045, best_f1=0.9024278515803941
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.046504661440849304
step: 10, loss: 0.04874103143811226
step: 20, loss: 0.18091417849063873
step: 30, loss: 0.039404451847076416
step: 40, loss: 0.06302233040332794
step: 50, loss: 0.003557836404070258
step: 60, loss: 0.005786739755421877
step: 70, loss: 0.002111660083755851
step: 80, loss: 0.0007198671228252351
step: 90, loss: 0.002726446371525526
step: 100, loss: 0.0038691298104822636
step: 110, loss: 0.0044686757028102875
step: 120, loss: 0.00828742329031229
step: 130, loss: 0.003885135753080249
step: 140, loss: 0.006907292176038027
step: 150, loss: 0.05477523431181908
step: 160, loss: 0.023478079587221146
step: 170, loss: 0.011349444277584553
step: 180, loss: 0.0013559970539063215
step: 190, loss: 0.002857985207810998
step: 200, loss: 0.05478991940617561
step: 210, loss: 0.04218299314379692
step: 220, loss: 0.0008949812618084252
step: 230, loss: 0.0023301616311073303
step: 240, loss: 0.031006738543510437
step: 250, loss: 0.0010051662102341652
```

```
step: 260, loss: 0.11639659106731415
step: 270, loss: 0.17777664959430695
step: 280, loss: 0.047467589378356934
step: 290, loss: 0.035955745726823381
step: 300, loss: 0.06544093042612076
step: 310, loss: 0.013608930632472038
step: 320, loss: 0.04850758984684944
step: 330, loss: 0.0071489629335701466
step: 340, loss: 0.008325188420712948
step: 350, loss: 0.019342826679348946
step: 360, loss: 0.03391164168715477
step: 370, loss: 0.010790492407977581
step: 380, loss: 0.05897977203130722
step: 390, loss: 0.006635407917201519
step: 400, loss: 0.04250386357307434
step: 410, loss: 0.011027957312762737
step: 420, loss: 0.0024728854186832905
step: 430, loss: 0.007059520110487938
step: 440, loss: 0.004716601222753525
step: 450, loss: 0.017784005030989647
step: 460, loss: 0.002381077501922846
step: 470, loss: 0.03749258071184158
step: 480, loss: 0.014416718855500221
step: 490, loss: 0.0129961296916008
step: 500, loss: 0.050927598029375076
step: 510, loss: 0.07143392413854599
step: 520, loss: 0.008467739447951317
step: 530, loss: 0.02069026604294777
/usr/local/lib/python3.7/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=0.9016853932584271, f1=0.9022695692450209,
best_f1=0.9024278515803941
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0168235432356596
step: 10, loss: 0.001130398246459663
step: 20, loss: 0.00276320893317461
step: 30, loss: 0.001820629695430398
step: 40, loss: 0.0392728820443153
step: 50, loss: 0.003742064582183957
step: 60, loss: 0.0032326197251677513
step: 70, loss: 0.062252126663412094
step: 80, loss: 0.0009310888708569109
step: 90, loss: 0.004502010531723499
step: 100, loss: 0.152687668800354
step: 110, loss: 0.041365355253219604
```

```
step: 120, loss: 0.014398454688489437
step: 130, loss: 0.049829576164484024
step: 140, loss: 0.0049765147268772125
step: 150, loss: 0.155941441655159
step: 160, loss: 0.0006902926252223551
step: 170, loss: 0.002332025207579136
step: 180, loss: 0.0037553380243480206
step: 190, loss: 0.011430523358285427
step: 200, loss: 0.0009303694241680205
step: 210, loss: 0.0006819806876592338
step: 220, loss: 0.0004009441181551665
step: 230, loss: 0.00024074710381682962
step: 240, loss: 0.013756124302744865
step: 250, loss: 0.0008452875190414488
step: 260, loss: 0.001350217149592936
step: 270, loss: 0.02362663671374321
step: 280, loss: 0.04301690682768822
step: 290, loss: 0.13955339789390564
step: 300, loss: 0.0013697873800992966
step: 310, loss: 0.13164499402046204
step: 320, loss: 0.2373807430267334
step: 330, loss: 0.03983522579073906
step: 340, loss: 0.05220731720328331
step: 350, loss: 0.18351693451404572
step: 360, loss: 0.012327308766543865
step: 370, loss: 0.013137539848685265
step: 380, loss: 0.005391756072640419
step: 390, loss: 0.05720507353544235
step: 400, loss: 0.000741739699151367
step: 410, loss: 0.003923739306628704
step: 420, loss: 0.1424456536769867
step: 430, loss: 0.03387947380542755
step: 440, loss: 0.04887174442410469
step: 450, loss: 0.004868742544203997
step: 460, loss: 0.18755395710468292
step: 470, loss: 0.03384498134255409
step: 480, loss: 0.004401968326419592
step: 490, loss: 0.0038866091053932905
step: 500, loss: 0.0035432344302535057
step: 510, loss: 0.0037005585618317127
step: 520, loss: 0.01767270267009735
step: 530, loss: 0.008196122944355011
/usr/local/lib/python3.7/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=0.8965839962564344, f1=0.8930581613508444,
best_f1=0.9024278515803941
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

```
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05818792060017586
step: 10, loss: 0.0067728739231824875
step: 20, loss: 0.001145325368270278
step: 30, loss: 0.16449011862277985
step: 40, loss: 0.0033482760190963745
step: 50, loss: 0.05491955950856209
step: 60, loss: 0.060293812304735184
step: 70, loss: 0.029794897884130478
step: 80, loss: 0.0009324475540779531
step: 90, loss: 0.01515840645879507
step: 100, loss: 0.021910659968852997
step: 110, loss: 0.0161609947681427
step: 120, loss: 0.0006409492343664169
step: 130, loss: 0.002629793481901288
step: 140, loss: 0.001489768153987825
step: 150, loss: 0.016152530908584595
step: 160, loss: 0.003234212752431631
step: 170, loss: 0.022240329533815384
step: 180, loss: 0.0007824453641660511
step: 190, loss: 0.0039294445887207985
step: 200, loss: 0.023793358355760574
step: 210, loss: 0.0021905461326241493
step: 220, loss: 0.0004272395162843168
step: 230, loss: 0.0014704412315040827
step: 240, loss: 0.004464120138436556
step: 250, loss: 0.02045774832367897
step: 260, loss: 0.00295657548122108
step: 270, loss: 0.00042493545333854854
step: 280, loss: 0.0047918627969920635
step: 290, loss: 0.0009183658985421062
step: 300, loss: 0.0013286909088492393
step: 310, loss: 0.030605768784880638
step: 320, loss: 0.026523951441049576
step: 330, loss: 0.025993570685386658
step: 340, loss: 0.016605094075202942
step: 350, loss: 0.03420144319534302
step: 360, loss: 0.030508195981383324
step: 370, loss: 0.0009181287023238838
step: 380, loss: 0.008777698501944542
step: 390, loss: 0.0007157435175031424
step: 400, loss: 0.001831302186474204
step: 410, loss: 0.000981678836978972
step: 420, loss: 0.0014307233504951
step: 430, loss: 0.005932639352977276
step: 440, loss: 0.0007798600709065795
step: 450, loss: 0.0005236382130533457
```

```
step: 460, loss: 0.0018915541004389524
step: 470, loss: 0.07400970906019211
step: 480, loss: 0.0015230578137561679
step: 490, loss: 0.03080444596707821
step: 500, loss: 0.00029781064949929714
step: 510, loss: 0.0016846958314999938
step: 520, loss: 0.07203815132379532
step: 530, loss: 0.0028157914057374
/usr/local/lib/python3.7/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=0.8944866920152091, f1=0.892018779342723,
best_f1=0.9024278515803941
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014443150721490383
step: 10, loss: 0.03480453044176102
step: 20, loss: 0.008545254357159138
step: 30, loss: 0.0015345766441896558
step: 40, loss: 0.002126599196344614
step: 50, loss: 0.0007447025272995234
step: 60, loss: 0.0019957188051193953
step: 70, loss: 0.0011813411256298423
step: 80, loss: 0.00015846274618525058
step: 90, loss: 0.0013563253451138735
step: 100, loss: 0.0024449494667351246
step: 110, loss: 0.0015349754830822349
step: 120, loss: 0.010929951444268227
step: 130, loss: 0.00031486086663790047
step: 140, loss: 0.00015251318109221756
step: 150, loss: 0.0010061472421512008
step: 160, loss: 0.0001758740982040763
step: 170, loss: 0.00022912138956598938
step: 180, loss: 0.002866990165784955
step: 190, loss: 0.003967075143009424
step: 200, loss: 0.0004000603221356869
step: 210, loss: 0.10978227108716965
step: 220, loss: 0.16444151103496552
step: 230, loss: 0.00046320160618051887
step: 240, loss: 0.0240371953696960125
step: 250, loss: 0.0003465019981376827
step: 260, loss: 0.0127321882173419
step: 270, loss: 0.0002557092811912298
step: 280, loss: 0.010750056244432926
step: 290, loss: 0.002118577715009451
step: 300, loss: 0.00021593578276224434
step: 310, loss: 0.0339195691049099
```

```
step: 320, loss: 0.00021848478354513645
step: 330, loss: 0.03319605439901352
step: 340, loss: 0.0006553017301484942
step: 350, loss: 0.0012728434521704912
step: 360, loss: 0.0003364357980899513
step: 370, loss: 0.0007357750437222421
step: 380, loss: 0.0008805306861177087
step: 390, loss: 0.00020213500829413533
step: 400, loss: 0.012506525963544846
step: 410, loss: 0.00022427826479543
step: 420, loss: 0.00033036895911209285
step: 430, loss: 0.0038676862604916096
step: 440, loss: 0.00016820601013023406
step: 450, loss: 0.004832401871681213
step: 460, loss: 0.004702270496636629
step: 470, loss: 0.0009455378749407828
step: 480, loss: 0.008849005214869976
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 490, loss: 0.008762999437749386
step: 500, loss: 0.0982901006937027
step: 510, loss: 0.009933454092592001
step: 520, loss: 0.03845953941345215
step: 530, loss: 0.0003758287930395454
/usr/local/lib/python3.7/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=0.8980156898938624, f1=0.9021043000914912,
best_f1=0.9024278515803941
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0024792344775505088
step: 10, loss: 0.00026334499125368893
step: 20, loss: 0.0009884879691526294
step: 30, loss: 0.0006957259611226618
step: 40, loss: 0.03308071941137314
step: 50, loss: 0.0013312960509210825
step: 60, loss: 0.0007207849994301796
step: 70, loss: 0.07521205395460129
step: 80, loss: 0.00013142221723683178
step: 90, loss: 0.0007901063654571772
step: 100, loss: 0.002476635854691267
step: 110, loss: 0.016840225085616112
step: 120, loss: 0.010282684117555618
step: 130, loss: 0.0006362953572534025
step: 140, loss: 0.0017567151226103306
step: 150, loss: 9.657489863457158e-05
step: 160, loss: 0.0002503430296201259
```

```
step: 170, loss: 0.00027715499163605273
step: 180, loss: 0.0015666650142520666
step: 190, loss: 0.00021685188403353095
step: 200, loss: 0.002851970959454775
step: 210, loss: 0.00029411280411295593
step: 220, loss: 0.0003996852319687605
step: 230, loss: 0.00022646604338660836
step: 240, loss: 0.0007052707951515913
step: 250, loss: 0.000277684215689078
step: 260, loss: 0.0007287433254532516
step: 270, loss: 0.0023264228366315365
step: 280, loss: 0.00039995406405068934
step: 290, loss: 0.0001313464599661529
step: 300, loss: 0.0014872155152261257
step: 310, loss: 0.0018397885141894221
step: 320, loss: 0.0007089424761943519
step: 330, loss: 0.0003161795320920646
step: 340, loss: 0.0003209565475117415
step: 350, loss: 0.000876928330399096
step: 360, loss: 0.023944811895489693
step: 370, loss: 0.002798652509227395
step: 380, loss: 0.0033345057163387537
step: 390, loss: 9.439962013857439e-05
step: 400, loss: 0.015366421081125736
step: 410, loss: 0.0005278222379274666
step: 420, loss: 0.0038459207862615585
step: 430, loss: 0.0002538116823416203
step: 440, loss: 0.02522345259785652
step: 450, loss: 0.0006146744708530605
step: 460, loss: 0.0024593474809080362
step: 470, loss: 0.0009777734521776438
step: 480, loss: 0.0010794757399708033
step: 490, loss: 0.0004258249537087977
step: 500, loss: 0.0015527355717495084
step: 510, loss: 0.0013758614659309387
step: 520, loss: 0.0004893289878964424
step: 530, loss: 0.00012623623479157686
/usr/local/lib/python3.7/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=0.8986742424242423, f1=0.8925541941564562,
best_f1=0.9024278515803941
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0028213008772581816
step: 10, loss: 0.0026698512956500053
step: 20, loss: 0.00010752307571237907
```

```
step: 30, loss: 0.00018058765272144228
step: 40, loss: 0.0008703155326656997
step: 50, loss: 0.00010895443119807169
step: 60, loss: 0.00046292212209664285
step: 70, loss: 0.0023441577795892954
step: 80, loss: 0.0032039149664342403
step: 90, loss: 0.00014104608271736652
step: 100, loss: 0.0008585829054936767
step: 110, loss: 0.0014072894118726254
step: 120, loss: 0.0016104384558275342
step: 130, loss: 0.016337135806679726
step: 140, loss: 0.00021666209799551226
step: 150, loss: 0.0070908102206885815
step: 160, loss: 0.0001935803738888353
step: 170, loss: 0.0009145893855020404
step: 180, loss: 0.023910900577902794
step: 190, loss: 0.0014169187052175403
step: 200, loss: 0.10000701993703842
step: 210, loss: 0.00022000166063662618
step: 220, loss: 0.00031259364914149046
step: 230, loss: 0.00011607375927269459
step: 240, loss: 0.00019566844275686689
step: 250, loss: 0.0010012281127274036
step: 260, loss: 0.00027600929024629295
step: 270, loss: 0.0207988228648901
step: 280, loss: 0.0015373709611594677
step: 290, loss: 0.0005813642055727541
step: 300, loss: 0.010267230682075024
step: 310, loss: 0.004650723189115524
step: 320, loss: 0.00012834934750571847
step: 330, loss: 0.0007007240201346576
step: 340, loss: 0.0002999606658704579
step: 350, loss: 0.00028699610265903175
step: 360, loss: 0.00022695887309964746
step: 370, loss: 0.0004866510280407965
step: 380, loss: 0.00025271569029428065
step: 390, loss: 0.000206931188586168
step: 400, loss: 0.00043678932706825435
step: 410, loss: 0.00035418971674516797
step: 420, loss: 0.02089591883122921
step: 430, loss: 0.0023190707433968782
step: 440, loss: 0.00012454162060748786
step: 450, loss: 0.002178195398300886
step: 460, loss: 0.0003653698367998004
step: 470, loss: 0.00018337472283747047
step: 480, loss: 0.0009468558127991855
step: 490, loss: 0.06607000529766083
step: 500, loss: 0.0003933984262403101
```

```
step: 510, loss: 0.00034950400004163384
step: 520, loss: 0.00015780962712597102
step: 530, loss: 0.00014475842181127518
/usr/local/lib/python3.7/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=0.8964200477326969, f1=0.8918406072106261,
best_f1=0.9024278515803941
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00380088877864182
step: 10, loss: 0.14035730063915253
step: 20, loss: 0.00010507521801628172
step: 30, loss: 0.0005361653747968376
step: 40, loss: 0.00014171478687785566
step: 50, loss: 0.00016301765572279692
step: 60, loss: 0.00015449011698365211
step: 70, loss: 0.00011102396820206195
step: 80, loss: 0.0001406080846209079
step: 90, loss: 0.00028567833942361176
step: 100, loss: 0.0015274633187800646
step: 110, loss: 0.0013511243741959333
step: 120, loss: 0.0008327129180543125
step: 130, loss: 0.000169809180079028
step: 140, loss: 0.0001683347945800051
step: 150, loss: 0.00025515753077343106
step: 160, loss: 0.019721725955605507
step: 170, loss: 0.00033354224609211087
step: 180, loss: 0.00014663039473816752
step: 190, loss: 0.00027093823882751167
step: 200, loss: 0.001635837135836482
step: 210, loss: 0.00026505423011258245
step: 220, loss: 0.0002216644206782803
step: 230, loss: 8.807596896076575e-05
step: 240, loss: 8.62732995301485e-05
step: 250, loss: 0.011143572628498077
step: 260, loss: 0.00010741022560978308
step: 270, loss: 0.0008634878322482109
step: 280, loss: 0.00046146559179760516
step: 290, loss: 0.0004729192878585309
step: 300, loss: 0.028328292071819305
step: 310, loss: 0.019365524873137474
step: 320, loss: 0.009465659037232399
step: 330, loss: 0.0022298935800790787
step: 340, loss: 0.000500315974932164
step: 350, loss: 0.0003309686726424843
step: 360, loss: 0.00025973585434257984
```

```
step: 370, loss: 0.0009342014091089368
step: 380, loss: 0.00013915609451942146
step: 390, loss: 0.0004793587722815573
step: 400, loss: 0.00010767983621917665
step: 410, loss: 0.00010515625035623088
step: 420, loss: 0.0031032392289489508
step: 430, loss: 0.022085681557655334
step: 440, loss: 0.00018466173787601292
step: 450, loss: 0.00016299067647196352
step: 460, loss: 0.0007141168462112546
step: 470, loss: 0.0006778523093089461
step: 480, loss: 0.0002176773559767753
step: 490, loss: 0.006305274553596973
step: 500, loss: 0.016902029514312744
step: 510, loss: 0.0011351616121828556
step: 520, loss: 0.00015673546295147139
step: 530, loss: 0.00010996011405950412
/usr/local/lib/python3.7/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=0.9015256588072121, f1=0.9020689655172414,
best_f1=0.9024278515803941
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002784742508083582
step: 10, loss: 0.00018937946879304945
step: 20, loss: 0.00013346440391615033
step: 30, loss: 0.00021259374625515193
step: 40, loss: 0.00017679555458016694
step: 50, loss: 0.00010586974531179294
step: 60, loss: 0.00020867571583944903
step: 70, loss: 0.00043241531238891184
step: 80, loss: 0.0004844185314141214
step: 90, loss: 0.011007490567862988
step: 100, loss: 6.87897700117901e-05
step: 110, loss: 0.002750572282820094
step: 120, loss: 0.00011477078805910423
step: 130, loss: 0.0011509439209476113
step: 140, loss: 0.00019466677762240395
step: 150, loss: 0.0007858124445192516
step: 160, loss: 0.000697922077961266
step: 170, loss: 0.0018411660566926003
step: 180, loss: 0.00010721656144596636
step: 190, loss: 7.452948921127245e-05
step: 200, loss: 9.447424963582307e-05
step: 210, loss: 0.0001569060405017808
step: 220, loss: 0.00010108183778356761
```

```
step: 230, loss: 0.08581320941448212
step: 240, loss: 0.00019935896852985024
step: 250, loss: 0.00777848344296217
step: 260, loss: 0.001059732399880886
step: 270, loss: 9.073683031601831e-05
step: 280, loss: 0.01516098715364933
step: 290, loss: 0.0029810448177158833
step: 300, loss: 0.0005015431670472026
step: 310, loss: 0.00041449526906944481
step: 320, loss: 0.00020361154747661203
step: 330, loss: 0.00016780242731329054
step: 340, loss: 0.07528873533010483
step: 350, loss: 0.004972649738192558
step: 360, loss: 0.00033741569495759904
step: 370, loss: 0.0005323590012267232
step: 380, loss: 0.00019817489373963326
step: 390, loss: 9.230933937942609e-05
step: 400, loss: 9.723346738610417e-05
step: 410, loss: 0.000539363652933389
step: 420, loss: 0.000581035332288593
step: 430, loss: 0.00017203990137204528
step: 440, loss: 0.0005642606993205845
step: 450, loss: 0.00018796247604768723
step: 460, loss: 0.051832303404808044
step: 470, loss: 0.00017760050832293928
step: 480, loss: 0.0004726300830952823
step: 490, loss: 0.0008024044800549746
step: 500, loss: 0.0031276303343474865
step: 510, loss: 0.00012532109394669533
step: 520, loss: 0.003923980984836817
step: 530, loss: 0.002912473399192095
/usr/local/lib/python3.7/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=0.8978622327790974, f1=0.898248935163275,
best_f1=0.9024278515803941
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00016188561858143657
step: 10, loss: 6.447353371186182e-05
step: 20, loss: 0.017986850813031197
step: 30, loss: 0.009709252044558525
step: 40, loss: 0.0001454859593650326
step: 50, loss: 0.00020883948309347034
step: 60, loss: 0.00011826863919850439
step: 70, loss: 0.00014086294686421752
step: 80, loss: 0.00018477799312677234
```

```
step: 90, loss: 0.00023997851531021297
step: 100, loss: 0.00011660567543003708
step: 110, loss: 0.0005238850135356188
step: 120, loss: 0.00026438466738909483
step: 130, loss: 0.00042018061503767967
step: 140, loss: 0.00012513634283095598
step: 150, loss: 0.0003412093501538038
step: 160, loss: 7.598313823109493e-05
step: 170, loss: 0.00016907825192902237
step: 180, loss: 0.00011907545558642596
step: 190, loss: 0.00018877522961702198
step: 200, loss: 0.00507873622700572
step: 210, loss: 0.0005150459473952651
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 220, loss: 0.00022354390239343047
step: 230, loss: 7.35078428988345e-05
step: 240, loss: 0.0002869732561521232
step: 250, loss: 0.08468134701251984
step: 260, loss: 0.00034940382465720177
step: 270, loss: 0.00013661156117450446
step: 280, loss: 2.7580916139413603e-05
step: 290, loss: 0.00010324132745154202
step: 300, loss: 0.00013586321438197047
step: 310, loss: 0.00014984951121732593
step: 320, loss: 8.145287574734539e-05
step: 330, loss: 0.00041898421477526426
step: 340, loss: 0.0001849143736762926
step: 350, loss: 0.0038655083626508713
step: 360, loss: 0.00011076035298174247
step: 370, loss: 0.0002883489360101521
step: 380, loss: 0.0003677412460092455
step: 390, loss: 5.6418786698486656e-05
step: 400, loss: 5.451968536362983e-05
step: 410, loss: 0.0001739011495374143
step: 420, loss: 6.452278466895223e-05
step: 430, loss: 5.534588490263559e-05
step: 440, loss: 8.834496111376211e-05
step: 450, loss: 8.952910138759762e-05
step: 460, loss: 3.9533591916551813e-05
step: 470, loss: 0.0004258534754626453
step: 480, loss: 0.00025354721583426
step: 490, loss: 0.00011631694360403344
step: 500, loss: 0.00020389797282405198
step: 510, loss: 7.612406625412405e-05
step: 520, loss: 3.566782470443286e-05
step: 530, loss: 7.817372534191236e-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=0.900612341026849, f1=0.9016853932584271,
best_f1=0.9024278515803941
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.501081174472347e-05
step: 10, loss: 7.783534965710714e-05
step: 20, loss: 0.00014949258184060454
step: 30, loss: 0.001724373665638268
step: 40, loss: 5.013326517655514e-05
step: 50, loss: 7.346543134190142e-05
step: 60, loss: 4.7229845222318545e-05
step: 70, loss: 0.0003402344882488251
step: 80, loss: 0.00011168492346769199
step: 90, loss: 3.54177100234665e-05
step: 100, loss: 8.189905202016234e-05
step: 110, loss: 0.00011201624147361144
step: 120, loss: 2.6242858439218253e-05
step: 130, loss: 4.581154644256458e-05
step: 140, loss: 0.0026101004332304
step: 150, loss: 9.565486834617332e-05
step: 160, loss: 0.0005078933318145573
step: 170, loss: 0.00016301768482662737
step: 180, loss: 5.453287667478435e-05
step: 190, loss: 7.68046811572276e-05
step: 200, loss: 8.349611016456038e-05
step: 210, loss: 0.00012982069165445864
step: 220, loss: 0.00019407921354286373
step: 230, loss: 0.0002587500202935189
step: 240, loss: 5.967416655039415e-05
step: 250, loss: 5.45035518415568e-05
step: 260, loss: 2.6876947231357917e-05
step: 270, loss: 0.0023879087530076504
step: 280, loss: 2.9655386242666282e-05
step: 290, loss: 0.0002333710581297055
step: 300, loss: 0.0005426006973721087
step: 310, loss: 0.00010585772542981431
step: 320, loss: 4.539620204013772e-05
step: 330, loss: 0.0005600936128757894
step: 340, loss: 0.0001366447249893099
step: 350, loss: 9.733738988870755e-05
step: 360, loss: 0.004700818099081516
step: 370, loss: 0.0014450266025960445
step: 380, loss: 5.7238001318182796e-05
step: 390, loss: 5.610305379377678e-05
step: 400, loss: 6.97882060194388e-05
step: 410, loss: 5.240006430540234e-05
```

```
step: 420, loss: 2.353966556256637e-05
step: 430, loss: 0.016926201060414314
step: 440, loss: 2.3841241272748448e-05
step: 450, loss: 0.0003056085843127221
step: 460, loss: 0.00011689451639540493
step: 470, loss: 0.0002714233414735645
step: 480, loss: 3.0102926757535897e-05
step: 490, loss: 3.385653690202162e-05
step: 500, loss: 4.951415758114308e-05
step: 510, loss: 4.448063555173576e-05
step: 520, loss: 0.00012954435078427196
step: 530, loss: 6.161248165881261e-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=0.9024856596558317, f1=0.89720511605874,
best_f1=0.9024278515803941
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.685102637973614e-05
step: 10, loss: 0.0001602011325303465
step: 20, loss: 0.00010176684736507013
step: 30, loss: 0.0002353454619878903
step: 40, loss: 0.00020817460608668625
step: 50, loss: 2.268640855618287e-05
step: 60, loss: 2.968461558339186e-05
step: 70, loss: 0.0001736302801873535
step: 80, loss: 5.792028969153762e-05
step: 90, loss: 5.962306386209093e-05
step: 100, loss: 3.278095755376853e-05
step: 110, loss: 9.891294757835567e-05
step: 120, loss: 0.0028403063770383596
step: 130, loss: 4.8351954319514334e-05
step: 140, loss: 4.4172968046041206e-05
step: 150, loss: 0.0002248253586035568
step: 160, loss: 0.005719916895031929
step: 170, loss: 0.00012977371807210147
step: 180, loss: 5.5685661209281534e-05
step: 190, loss: 8.19832639535889e-05
step: 200, loss: 3.079211091971956e-05
step: 210, loss: 3.714976628543809e-05
step: 220, loss: 2.4917579139582813e-05
step: 230, loss: 3.356681554578245e-05
step: 240, loss: 3.917713911505416e-05
step: 250, loss: 0.00034198956564068794
step: 260, loss: 0.0004955343902111053
step: 270, loss: 0.00525476410984993
```

```
step: 280, loss: 3.518002631608397e-05
step: 290, loss: 0.005738688632845879
step: 300, loss: 0.07695567607879639
step: 310, loss: 0.0001761810708558187
step: 320, loss: 7.125394040485844e-05
step: 330, loss: 0.0026933467015624046
step: 340, loss: 6.151381239760667e-05
step: 350, loss: 0.00014177252887748182
step: 360, loss: 5.811472874484025e-05
step: 370, loss: 2.3997681637411006e-05
step: 380, loss: 7.092361192917451e-05
step: 390, loss: 0.00011323960643494502
step: 400, loss: 2.6664618417271413e-05
step: 410, loss: 0.00012239639181643724
step: 420, loss: 8.055374200921506e-05
step: 430, loss: 2.87545844912529e-05
step: 440, loss: 0.11375642567873001
step: 450, loss: 0.0016367629868909717
step: 460, loss: 2.2187394279171713e-05
step: 470, loss: 6.999156175879762e-05
step: 480, loss: 7.651443593204021e-05
step: 490, loss: 7.246466702781618e-05
step: 500, loss: 8.713372517377138e-05
step: 510, loss: 4.770417945110239e-05
step: 520, loss: 3.810680209426209e-05
step: 530, loss: 5.3352803661255166e-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=0.9038642789820923, f1=0.9016853932584271,
best_f1=0.9024278515803941
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.")
5742it [00:23, 241.57it/s]
load_f1 = 0.9053803339517625
real_f1 = 0.9033749422098938
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 [00:16, 259.80it/s]
```

1.1.4 Amazon-Google - Running the matcher

```python
[ ]: # 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 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: 0.8456778526306152
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.06456080079078674

66
```

```
step: 20, loss: 0.3807266056537628
step: 30, loss: 0.37620556354522705
step: 40, loss: 0.5145447850227356
step: 50, loss: 0.30867263674736023
step: 60, loss: 0.3658502697944641
step: 70, loss: 0.27315881848335266
step: 80, loss: 0.2973671555519104
step: 90, loss: 0.4409922659397125
step: 100, loss: 0.20932962000370026
step: 110, loss: 0.3095425069332123
step: 120, loss: 0.2611664831638336
step: 130, loss: 0.22988401353359222
step: 140, loss: 0.3222893476486206
step: 150, loss: 0.3270992338657379
step: 160, loss: 0.26680129766464233
step: 170, loss: 0.20563796162605286
step: 180, loss: 0.19600629806518555
step: 190, loss: 0.25065919756889343
step: 200, loss: 0.17640426754951477
step: 210, loss: 0.47964951395988464
/usr/local/lib/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=0.38080495356037153, f1=0.384, best_f1=0.384
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08926306664943695
step: 10, loss: 0.09255017340183258
step: 20, loss: 0.32549771666526794
step: 30, loss: 0.19920581579208374
step: 40, loss: 0.04807789623737335
step: 50, loss: 0.21108008921146393
step: 60, loss: 0.15374372899532318
step: 70, loss: 0.2710377871990204
step: 80, loss: 0.29782232642173767
step: 90, loss: 0.1975618153810501
step: 100, loss: 0.15362529456615448
step: 110, loss: 0.06938175112009048
step: 120, loss: 0.2047736793756485
step: 130, loss: 0.2615467309951782
step: 140, loss: 0.29125550389289856
step: 150, loss: 0.32056936621665955
step: 160, loss: 0.14926829934120178
```

```
step: 170, loss: 0.2144795060157776
step: 180, loss: 0.40540218353271484
step: 190, loss: 0.23824672400951385
step: 200, loss: 0.2477351874113083
step: 210, loss: 0.33392563462257385
/usr/local/lib/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=0.46843853820598, f1=0.499194847020934,
best_f1=0.499194847020934
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2814784348011017
step: 10, loss: 0.2609745264053345
step: 20, loss: 0.3620871901512146
step: 30, loss: 0.08418212085962296
step: 40, loss: 0.18132451176643372
step: 50, loss: 0.24412070214748383
step: 60, loss: 0.30269986391067505
step: 70, loss: 0.2088860124349594
step: 80, loss: 0.15277045965194702
step: 90, loss: 0.12222922593355179
step: 100, loss: 0.1791199892759323
step: 110, loss: 0.2711721956729889
step: 120, loss: 0.2230774611234665
step: 130, loss: 0.25552916526794434
step: 140, loss: 0.27694422006607056
step: 150, loss: 0.3136068284511566
step: 160, loss: 0.14250913262367249
step: 170, loss: 0.2360607236623764
step: 180, loss: 0.11848119646310806
step: 190, loss: 0.277439683675766
step: 200, loss: 0.2322743833065033
step: 210, loss: 0.30381691455841064
/usr/local/lib/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=0.46855345911949686, f1=0.49841269841269836,
best_f1=0.49841269841269836
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3300068974494934
step: 10, loss: 0.10503236204385757
step: 20, loss: 0.14540253579616547
step: 30, loss: 0.08137619495391846
step: 40, loss: 0.1437663733959198
step: 50, loss: 0.11366615444421768
step: 60, loss: 0.04640205577015877
step: 70, loss: 0.22081239521503448
step: 80, loss: 0.24357779324054718
step: 90, loss: 0.09490343928337097
step: 100, loss: 0.1942514330148697
step: 110, loss: 0.1484074741601944
step: 120, loss: 0.07102569192647934
step: 130, loss: 0.29381638765335083
step: 140, loss: 0.16887854039669037
step: 150, loss: 0.23458515107631683
step: 160, loss: 0.1458277851343155
step: 170, loss: 0.07249008119106293
step: 180, loss: 0.1816936880350113
step: 190, loss: 0.2502930462360382
step: 200, loss: 0.28147944808006287
step: 210, loss: 0.03510282561182976
/usr/local/lib/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=0.481981981981982, f1=0.48660714285714285,
best_f1=0.48660714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09829438477754593
step: 10, loss: 0.04703821241855621
step: 20, loss: 0.029856901615858078
step: 30, loss: 0.1271059364080429
step: 40, loss: 0.1289234459400177
step: 50, loss: 0.165321946144104
step: 60, loss: 0.11647560447454453
step: 70, loss: 0.15146994590759277
step: 80, loss: 0.12310336530208588
step: 90, loss: 0.040792711079120636
step: 100, loss: 0.09536556154489517
```

```
step: 110, loss: 0.01761579141020775
step: 120, loss: 0.04784056171774864
step: 130, loss: 0.15993095934391022
step: 140, loss: 0.08828330785036087
step: 150, loss: 0.07371941953897476
step: 160, loss: 0.10314375907182693
step: 170, loss: 0.09207654744386673
step: 180, loss: 0.2669273316860199
step: 190, loss: 0.06527397781610489
step: 200, loss: 0.10394404083490372
step: 210, loss: 0.09904501587152481
/usr/local/lib/python3.7/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=0.46208112874779544, f1=0.48359240069084625,
best_f1=0.48660714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10854185372591019
step: 10, loss: 0.037695176899433136
step: 20, loss: 0.07200755178928375
step: 30, loss: 0.18415983021259308
step: 40, loss: 0.030093995854258537
step: 50, loss: 0.15817750990390778
step: 60, loss: 0.21112015843391418
step: 70, loss: 0.02170580066740513
step: 80, loss: 0.07810580730438232
step: 90, loss: 0.07422081381082535
step: 100, loss: 0.01752845011651516
step: 110, loss: 0.029773229733109474
step: 120, loss: 0.028803814202547073
step: 130, loss: 0.01676318794488907
step: 140, loss: 0.11439143121242523
step: 150, loss: 0.09794951230287552
step: 160, loss: 0.34386181831359863
step: 170, loss: 0.08142910897731781
step: 180, loss: 0.01002525445073843
step: 190, loss: 0.12404526770114899
step: 200, loss: 0.020253516733646393
step: 210, loss: 0.038497425615787506
/usr/local/lib/python3.7/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=0.4549763033175355, f1=0.44028103044496486,
best_f1=0.48660714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.040582265704870224
step: 10, loss: 0.05900140479207039
step: 20, loss: 0.1373848021030426
step: 30, loss: 0.026689302176237106
step: 40, loss: 0.04383700340986252
step: 50, loss: 0.09173527359962463
step: 60, loss: 0.22432342171669006
step: 70, loss: 0.012117201462388039
step: 80, loss: 0.027664005756378174
step: 90, loss: 0.044858258217573166
step: 100, loss: 0.027407541871070862
step: 110, loss: 0.08844266831874847
step: 120, loss: 0.10744810849428177
step: 130, loss: 0.04327138140797615
step: 140, loss: 0.0036240662448108196
step: 150, loss: 0.08861348778009415
step: 160, loss: 0.03239312022924423
step: 170, loss: 0.021831003949046135
step: 180, loss: 0.013467919081449509
step: 190, loss: 0.03650767728686333
step: 200, loss: 0.042513057589530945
step: 210, loss: 0.11368042230606079
/usr/local/lib/python3.7/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=0.4365482233502538, f1=0.46942148760330576,
best_f1=0.48660714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0540497750043869
step: 10, loss: 0.04175443574786186
step: 20, loss: 0.057762958109378815
step: 30, loss: 0.004270041361451149
step: 40, loss: 0.04820191487669945
step: 50, loss: 0.06554319709539413
step: 60, loss: 0.22499170899391174
step: 70, loss: 0.03514831140637398
step: 80, loss: 0.03498339653015137
step: 90, loss: 0.03033401444554329
step: 100, loss: 0.00880111288279295
step: 110, loss: 0.004754809662699699
step: 120, loss: 0.15549759566783905
step: 130, loss: 0.06356823444366455
step: 140, loss: 0.03484221175312996
step: 150, loss: 0.008007647469639778
step: 160, loss: 0.09496346116065979
```

```
step: 170, loss: 0.05025826394557953
step: 180, loss: 0.07667262852191925
step: 190, loss: 0.032417938113212585
step: 200, loss: 0.05534468963742256
step: 210, loss: 0.12930867075920105
/usr/local/lib/python3.7/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=0.4383561643835616, f1=0.45889101338432126,
best_f1=0.48660714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.025095460936427116
step: 10, loss: 0.060237541794776917
step: 20, loss: 0.05865410342812538
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 30, loss: 0.025891628116369247
step: 40, loss: 0.07804465293884277
step: 50, loss: 0.03984227776527405
step: 60, loss: 0.12377537041902542
step: 70, loss: 0.006862349808216095
step: 80, loss: 0.028946902602910995
step: 90, loss: 0.10811996459960938
step: 100, loss: 0.018027685582637787
step: 110, loss: 0.010406683199107647
step: 120, loss: 0.1355234682559967
step: 130, loss: 0.07205598801374435
step: 140, loss: 0.06764546036720276
step: 150, loss: 0.183905690908432
step: 160, loss: 0.006033942569047213
step: 170, loss: 0.020103242248296738
step: 180, loss: 0.011365673504769802
step: 190, loss: 0.15257994830608368
step: 200, loss: 0.14937348663806915
step: 210, loss: 0.058171067386865616
/usr/local/lib/python3.7/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=0.42043222003929276, f1=0.4518664047151277,
best_f1=0.48660714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.011562732979655266
step: 10, loss: 0.05792880803346634
step: 20, loss: 0.011871728114783764
step: 30, loss: 0.006408113520592451
```

```
step: 40, loss: 0.008177890442311764
step: 50, loss: 0.01002572663128376
step: 60, loss: 0.061425354331731796
step: 70, loss: 0.013072936795651913
step: 80, loss: 0.0005557503900490701
step: 90, loss: 0.029567021876573563
step: 100, loss: 0.004337146412581205
step: 110, loss: 0.010704852640628815
step: 120, loss: 0.0031924883369356394
step: 130, loss: 0.07992718368768692
step: 140, loss: 0.009591088630259037
step: 150, loss: 0.08930835127830505
step: 160, loss: 0.08024940639734268
step: 170, loss: 0.019300252199172974
step: 180, loss: 0.11412985622882843
step: 190, loss: 0.12218368053436279
step: 200, loss: 0.009370430372655392
step: 210, loss: 0.008378822356462479
/usr/local/lib/python3.7/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=0.42750929368029744, f1=0.46099290780141844,
best_f1=0.48660714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0037921147886663675
step: 10, loss: 0.010303989052772522
step: 20, loss: 0.02413698472082615
step: 30, loss: 0.024672802537679672
step: 40, loss: 0.11547865718603134
step: 50, loss: 0.008998638018965721
step: 60, loss: 0.09179141372442245
step: 70, loss: 0.005306830629706383
step: 80, loss: 0.07145848125219345
step: 90, loss: 0.018952030688524246
step: 100, loss: 0.04799982160329819
step: 110, loss: 0.0264721792191267
step: 120, loss: 0.01034195814281702
step: 130, loss: 0.08392032235860825
step: 140, loss: 0.003943071234971285
step: 150, loss: 0.11766071617603302
step: 160, loss: 0.0020087396260350943
step: 170, loss: 0.06830031424760818
step: 180, loss: 0.004298518877476454
step: 190, loss: 0.16259972751140594
step: 200, loss: 0.0027929130010306835
step: 210, loss: 0.00494022062048316
```

```
/usr/local/lib/python3.7/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=0.4355400696864112, f1=0.46982055464926586,
best_f1=0.48660714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.013550794683396816
step: 10, loss: 0.005034329369664192
step: 20, loss: 0.019061675295233727
step: 30, loss: 0.002327112713828683
step: 40, loss: 0.23546244204044342
step: 50, loss: 0.0712028220295906
step: 60, loss: 0.009803787805140018
step: 70, loss: 0.005593783222138882
step: 80, loss: 0.01123836264014244
step: 90, loss: 0.0013146625133231282
step: 100, loss: 0.006586746778339148
step: 110, loss: 0.004865326918661594
step: 120, loss: 0.1389550417661667
step: 130, loss: 0.0017231358215212822
step: 140, loss: 0.0026501782704144716
step: 150, loss: 0.0009567174129188061
step: 160, loss: 0.015301130712032318
step: 170, loss: 0.0029126908630132675
step: 180, loss: 0.005071030929684639
step: 190, loss: 0.023630931973457336
step: 200, loss: 0.030125709250569344
step: 210, loss: 0.007196256425231695
/usr/local/lib/python3.7/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=0.3943089430894309, f1=0.43951612903225806,
best_f1=0.48660714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00034838836290873587
step: 10, loss: 0.005936694797128439
step: 20, loss: 0.006824334617704153
step: 30, loss: 0.009562456049025059
step: 40, loss: 0.0039743161760270596
step: 50, loss: 0.007665056735277176
step: 60, loss: 0.03089250437915325
step: 70, loss: 0.18680867552757263
step: 80, loss: 0.09470812976360321
step: 90, loss: 0.010552536696195602
```

```
step: 100, loss: 0.10676909238100052
step: 110, loss: 0.0027115463744848967
step: 120, loss: 0.01631317473948002
step: 130, loss: 0.00716894818469882
step: 140, loss: 0.0031693389173597097
step: 150, loss: 0.013568056747317314
step: 160, loss: 0.050691328942775726
step: 170, loss: 0.09254306554794312
step: 180, loss: 0.08573199808597565
step: 190, loss: 0.01208547130227089
step: 200, loss: 0.004055853933095932
step: 210, loss: 0.0051037282682955265
/usr/local/lib/python3.7/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=0.43214285714285716, f1=0.4667802385008518,
best_f1=0.48660714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007809730246663094
step: 10, loss: 0.023523647338151932
step: 20, loss: 0.15476615726947784
step: 30, loss: 0.1923884153366089
step: 40, loss: 0.0065032378770411015
step: 50, loss: 0.004020404070615768
step: 60, loss: 0.07262475788593292
step: 70, loss: 0.040449097752571106
step: 80, loss: 0.0840236023068428
step: 90, loss: 0.03414108604192734
step: 100, loss: 0.0032076742500066757
step: 110, loss: 0.003777500009164214
step: 120, loss: 0.005967878736555576
step: 130, loss: 0.0029075562488287687
step: 140, loss: 0.0014018932124599814
step: 150, loss: 0.04838516563177109
step: 160, loss: 0.03174687176942825
step: 170, loss: 0.07138338685035706
step: 180, loss: 0.001905735582113266
step: 190, loss: 0.0020229024812579155
step: 200, loss: 0.004740070551633835
step: 210, loss: 0.003920606803148985
/usr/local/lib/python3.7/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=0.4140625, f1=0.45112781954887216, best_f1=0.48660714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008598471991717815
step: 10, loss: 0.005418424028903246
step: 20, loss: 0.003756130114197731
step: 30, loss: 0.0009078930597752333
step: 40, loss: 0.005282688420265913
step: 50, loss: 0.0010650543263182044
step: 60, loss: 0.00431957608088851
step: 70, loss: 0.0037233878392726183
step: 80, loss: 0.0018068895442411304
step: 90, loss: 0.0014635270927101374
step: 100, loss: 0.0038453424349427223
step: 110, loss: 0.0035704656038433313
step: 120, loss: 0.055393021553754807
step: 130, loss: 0.05602944269776344
step: 140, loss: 0.010321719571948051
step: 150, loss: 0.007063591852784157
step: 160, loss: 0.006331038195639849
step: 170, loss: 0.023130172863602638
step: 180, loss: 0.01633705571293831
step: 190, loss: 0.015955591574311256
step: 200, loss: 0.013965553604066372
step: 210, loss: 0.09872263669967651
/usr/local/lib/python3.7/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=0.41221374045801534, f1=0.4618249534450652,
best_f1=0.48660714285714285
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.")
2293it [00:06, 329.58it/s]
load_f1 = 0.477751756440281
real_f1 = 0.46976744186046515
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, 253.34it/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: 0.8678837418556213
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.16431409120559692
step: 20, loss: 0.1421486884355545
step: 30, loss: 0.5015112161636353
step: 40, loss: 0.250747412443161

```
step: 50, loss: 0.3105010390281677
step: 60, loss: 0.36676767468452454
step: 70, loss: 0.17647258937358856
step: 80, loss: 0.5167737007141113
step: 90, loss: 0.24180659651756287
step: 100, loss: 0.22269174456596375
step: 110, loss: 0.24042633175849915
step: 120, loss: 0.4231984317302704
step: 130, loss: 0.3465990424156189
step: 140, loss: 0.33385542035102844
step: 150, loss: 0.26747599244117737
step: 160, loss: 0.2063254714012146
step: 170, loss: 0.40739062428474426
step: 180, loss: 0.2623629868030548
step: 190, loss: 0.19596511125564575
/usr/local/lib/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=0.4779220779220779, f1=0.4934383202099738,
best_f1=0.4934383202099738
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2633635699748993
step: 10, loss: 0.04786764085292816
step: 20, loss: 0.07715590298175812
step: 30, loss: 0.1611405313014984
step: 40, loss: 0.4265322685241699
step: 50, loss: 0.3493155539035797
step: 60, loss: 0.1917356699705124
step: 70, loss: 0.18155165016651154
step: 80, loss: 0.18772318959236145
step: 90, loss: 0.11851777136325836
step: 100, loss: 0.173895001411438
step: 110, loss: 0.15665972232818604
step: 120, loss: 0.27824679017066956
step: 130, loss: 0.25331220030784607
step: 140, loss: 0.3050307631492615
step: 150, loss: 0.08802862465381622
step: 160, loss: 0.04239024594426155
step: 170, loss: 0.21196013689041138
step: 180, loss: 0.1010487973690033
step: 190, loss: 0.3062446415424347
/usr/local/lib/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=0.6807387862796834, f1=0.6938775510204082,
best_f1=0.6938775510204082
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.14338579773902893
step: 10, loss: 0.17710936069488525
step: 20, loss: 0.0349930115044117
step: 30, loss: 0.08764366060495377
step: 40, loss: 0.044273633509874344
step: 50, loss: 0.10863310098648071
step: 60, loss: 0.03498607128858566
step: 70, loss: 0.22212651371955872
step: 80, loss: 0.09280115365982056
step: 90, loss: 0.05270369350910187
step: 100, loss: 0.11386103183031082
step: 110, loss: 0.2517056465148926
step: 120, loss: 0.08085281401872635
step: 130, loss: 0.022838188335299492
step: 140, loss: 0.06517168134450912
step: 150, loss: 0.14666874706745148
step: 160, loss: 0.14543737471103668
step: 170, loss: 0.057392898947000504
step: 180, loss: 0.13371647894382477
step: 190, loss: 0.20157945156097412
/usr/local/lib/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=0.7295285359801489, f1=0.7189873417721518,
best_f1=0.7189873417721518
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.12970228493213654
step: 10, loss: 0.0698310062289238
step: 20, loss: 0.015979383140802383
step: 30, loss: 0.059415753930807114
step: 40, loss: 0.013991555199027061
```

```
step: 50, loss: 0.07451887428760529
step: 60, loss: 0.07428291440010071
step: 70, loss: 0.04724973440170288
step: 80, loss: 0.028846148401498795
step: 90, loss: 0.06532188504934311
step: 100, loss: 0.04442692548036575
step: 110, loss: 0.052253950387239456
step: 120, loss: 0.028280247002840042
step: 130, loss: 0.25218021869659424
step: 140, loss: 0.06400696933269501
step: 150, loss: 0.00625191256403923
step: 160, loss: 0.043552033603191376
step: 170, loss: 0.024665849283337593
step: 180, loss: 0.1846565455198288
step: 190, loss: 0.05218977481126785
/usr/local/lib/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=0.7738693467336683, f1=0.7531806615776082,
best_f1=0.7531806615776082
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02527630887925625
step: 10, loss: 0.05514708161354065
step: 20, loss: 0.06019602715969086
step: 30, loss: 0.011832596734166145
step: 40, loss: 0.05788213759660721
step: 50, loss: 0.049737196415662766
step: 60, loss: 0.04365835338830948
step: 70, loss: 0.007714762352406979
step: 80, loss: 0.0766611248254776
step: 90, loss: 0.006482226308435202
step: 100, loss: 0.2147664576768875
step: 110, loss: 0.021052632480859756
step: 120, loss: 0.002196545246988535
step: 130, loss: 0.00985052902251482
step: 140, loss: 0.00418392289429903
step: 150, loss: 0.010351603850722313
step: 160, loss: 0.021520299836993217
step: 170, loss: 0.05964547023177147
step: 180, loss: 0.1136445477604866
step: 190, loss: 0.25972381234169006
/usr/local/lib/python3.7/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=0.7671957671957672, f1=0.7559055118110236,
best_f1=0.7531806615776082
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0232473127543392624
step: 10, loss: 0.001750568044371903
step: 20, loss: 0.03852487727999687
step: 30, loss: 0.07112712413072586
step: 40, loss: 0.09821425378322601
step: 50, loss: 0.008450164459645748
step: 60, loss: 0.019604306668043137
step: 70, loss: 0.1486021876335144
step: 80, loss: 0.0076541113667190075
step: 90, loss: 0.005934102926403284
step: 100, loss: 0.008492558263242245
step: 110, loss: 0.238446444272995
step: 120, loss: 0.012405682355165482
step: 130, loss: 0.007572017144411802
step: 140, loss: 0.05803164839744568
step: 150, loss: 0.008671501651406288
step: 160, loss: 0.011198555119335651
step: 170, loss: 0.0016735391691327095
step: 180, loss: 0.07497107982635498
step: 190, loss: 0.024338332936167717
/usr/local/lib/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=0.7807486631016043, f1=0.7564766839378239,
best_f1=0.7564766839378239
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010573090054094791
step: 10, loss: 0.04706812649965286
step: 20, loss: 0.006131439004093409
step: 30, loss: 0.08082418143749237
step: 40, loss: 0.0407940149307251
step: 50, loss: 0.029087474569678307
step: 60, loss: 0.01677493378520012
step: 70, loss: 0.0016355854459106922
step: 80, loss: 0.020633600652217865
```

```
step: 90, loss: 0.04044128954410553
step: 100, loss: 0.02509012073278427
step: 110, loss: 0.007132368162274361
step: 120, loss: 0.008940208703279495
step: 130, loss: 0.15429741144180298
step: 140, loss: 0.015190189704298973
step: 150, loss: 0.02633426897227764
step: 160, loss: 0.029835378751158714
step: 170, loss: 0.017054220661520958
step: 180, loss: 0.009696519933640957
step: 190, loss: 0.0445319227874279
/usr/local/lib/python3.7/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=0.7626262626262625,
best_f1=0.7564766839378239
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03261570259928703
step: 10, loss: 0.002620183164253831
step: 20, loss: 0.011599949561059475
step: 30, loss: 0.002687143161892891
step: 40, loss: 0.016655683517456055
step: 50, loss: 0.002307336777448654
step: 60, loss: 0.0064957174472510815
step: 70, loss: 0.0021175346337258816
step: 80, loss: 0.026682322844862938
step: 90, loss: 0.005398321896791458
step: 100, loss: 0.04491671547293663
step: 110, loss: 0.013606049120426178
step: 120, loss: 0.003257026895880699
step: 130, loss: 0.019658531993627548
step: 140, loss: 0.002923742402344942
step: 150, loss: 0.003060900606215
step: 160, loss: 0.1262994408607483
step: 170, loss: 0.005565932486206293
step: 180, loss: 0.06281828880310059
step: 190, loss: 0.17441576719284058
/usr/local/lib/python3.7/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=0.768, f1=0.7407407407407407, best_f1=0.7564766839378239
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005637850612401962
step: 10, loss: 0.05000387504696846
```

```
step: 20, loss: 0.031668633222579956
step: 30, loss: 0.008642280474305153
step: 40, loss: 0.011809336952865124
step: 50, loss: 0.019147520884871483
step: 60, loss: 0.0026575883384793997
step: 70, loss: 0.007522759959101677
step: 80, loss: 0.003777993842959404
step: 90, loss: 0.06488318741321564
step: 100, loss: 0.0026136483065783978
step: 110, loss: 0.00195132486987859
step: 120, loss: 0.06728202849626541
step: 130, loss: 0.0027576705906540155
step: 140, loss: 0.004403950180858374
step: 150, loss: 0.03507433086633682
step: 160, loss: 0.0016380075830966234
step: 170, loss: 0.035787977278232574
step: 180, loss: 0.027322161942720413
step: 190, loss: 0.022305095568299294
/usr/local/lib/python3.7/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=0.7440633245382585, f1=0.7180851063829788,
best_f1=0.7564766839378239
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002110648900270462
step: 10, loss: 0.01807994395494461
step: 20, loss: 0.0022865834180265665
step: 30, loss: 0.007127002347260714
step: 40, loss: 0.003843077691271901
step: 50, loss: 0.009056451730430126
step: 60, loss: 0.011387858539819717
step: 70, loss: 0.00794499833136797
step: 80, loss: 0.0006221596850082278
step: 90, loss: 0.008197651244699955
step: 100, loss: 0.0058294604532420635
step: 110, loss: 0.07864286750555038
step: 120, loss: 0.0009001166326925159
step: 130, loss: 0.0005471630720421672
step: 140, loss: 0.0018644073279574513
step: 150, loss: 0.0037562339566648006
step: 160, loss: 0.4237498342990875
step: 170, loss: 0.0551852323114872
step: 180, loss: 0.04834866151213646
step: 190, loss: 0.023153753951191902
/usr/local/lib/python3.7/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=0.7637362637362638, f1=0.7444444444444444,
best_f1=0.7564766839378239
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014430033974349499
step: 10, loss: 0.007985611446201801
step: 20, loss: 0.005336422473192215
step: 30, loss: 0.031627390533685684
step: 40, loss: 0.005614247638732195
step: 50, loss: 0.006420521065592766
step: 60, loss: 0.0016789919463917613
step: 70, loss: 0.0006192016880959272
step: 80, loss: 0.0012662721564993262
step: 90, loss: 0.0010186316212639213
step: 100, loss: 0.0014447278808802366
step: 110, loss: 0.0009033339447341859
step: 120, loss: 0.007769385352730751
step: 130, loss: 0.002590509131550789
step: 140, loss: 0.003533207578584552
step: 150, loss: 0.0022688566241413355
step: 160, loss: 0.002308236202225089
step: 170, loss: 0.06975770741701126
step: 180, loss: 0.023638278245925903
step: 190, loss: 0.0017413630848750472
/usr/local/lib/python3.7/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=0.7594936708860759, f1=0.7272727272727272,
best_f1=0.7564766839378239
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005470136180520058
step: 10, loss: 0.02031565085053444
step: 20, loss: 0.0024342695251107216
step: 30, loss: 0.0008079961407929659
step: 40, loss: 0.0009177503525279462
step: 50, loss: 0.020809095352888107
step: 60, loss: 0.0004675366508308798
step: 70, loss: 0.0035918226931244135
step: 80, loss: 0.0060506658628582954
step: 90, loss: 0.0024091608356684446
step: 100, loss: 0.000633000920061022
step: 110, loss: 0.0003271695750299841
step: 120, loss: 0.0019564705435186625
step: 130, loss: 0.0007202517008408904
```

```
step: 140, loss: 0.0028176107443869114
step: 150, loss: 0.0017312112031504512
step: 160, loss: 0.032560329884290695
step: 170, loss: 0.01804855465888977
step: 180, loss: 0.0008294570725411177
step: 190, loss: 0.009940053336322308
/usr/local/lib/python3.7/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=0.769230769230769, f1=0.7553191489361701,
best_f1=0.7564766839378239
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004076894838362932
step: 10, loss: 0.007139982655644417
step: 20, loss: 0.0019623765256255865
step: 30, loss: 0.06703481823205948
step: 40, loss: 0.00020507631415966898
step: 50, loss: 0.005515545140951872
step: 60, loss: 0.0017408673884347081
step: 70, loss: 0.002031204756349325
step: 80, loss: 0.017206432297825813
step: 90, loss: 0.006660028826445341
step: 100, loss: 0.0010904045775532722
step: 110, loss: 0.0021526122000068426
step: 120, loss: 0.001422008266672492
step: 130, loss: 0.042837128043174744
step: 140, loss: 0.0005386951379477978
step: 150, loss: 0.06873048096895218
step: 160, loss: 0.0006219229544512928
step: 170, loss: 0.00047481898218393326
step: 180, loss: 0.004974374081939459
step: 190, loss: 0.0027664960362017155
/usr/local/lib/python3.7/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=0.7578947368421052, f1=0.7546174142480211,
best_f1=0.7564766839378239
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.013842931017279625
step: 10, loss: 0.001934035331942141
step: 20, loss: 0.013724755495786667
step: 30, loss: 0.005519203841686249
step: 40, loss: 0.000736475340090692
step: 50, loss: 0.004557224456220865
```

```
step: 60, loss: 0.044383157044649124
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 70, loss: 0.12283571809530258
step: 80, loss: 0.0007859551114961505
step: 90, loss: 0.030091840773820877
step: 100, loss: 0.0024280776269733906
step: 110, loss: 0.09973283112049103
step: 120, loss: 0.0034387842752039433
step: 130, loss: 0.00034097698517143726
step: 140, loss: 0.0004563444817904383
step: 150, loss: 0.0007476042956113815
step: 160, loss: 0.018207263201475143
step: 170, loss: 0.14001844823360443
step: 180, loss: 0.0027432686183601618
step: 190, loss: 0.0005339630879461765
/usr/local/lib/python3.7/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=0.7563451776649746, f1=0.732824427480916,
best_f1=0.7564766839378239
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006354871671646833
step: 10, loss: 0.000927323242649436
step: 20, loss: 0.00091799336951226
step: 30, loss: 0.0015073786489665508
step: 40, loss: 0.002638385398313403
step: 50, loss: 0.014604699797928333
step: 60, loss: 0.08576289564371109
step: 70, loss: 0.0010523569071665406
step: 80, loss: 0.0007302048616111279
step: 90, loss: 0.025354884564876556
step: 100, loss: 0.0018072273815050721
step: 110, loss: 0.0011934958165511489
step: 120, loss: 0.008690353482961655
step: 130, loss: 0.0008788208360783756
step: 140, loss: 0.016553644090890884
step: 150, loss: 0.0012686122208833694
step: 160, loss: 0.0018825225997716188
step: 170, loss: 0.0006593374419026077
step: 180, loss: 0.0021079801954329014
step: 190, loss: 0.008673657663166523
/usr/local/lib/python3.7/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=0.7639257294429709, f1=0.7580645161290323,
best_f1=0.7564766839378239
```

```
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.")
2049it [00:09, 223.00it/s]
load_f1 = 0.7169811320754718
real_f1 = 0.7200000000000001
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, 252.05it/s]
```

1.2 DA TEXTUAL

1.2.1 Abt-Buy - Running the matcher

```python
[ ]: # 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 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: 0.8606671094894409
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.21967338025569916
step: 20, loss: 0.14477181434631348
step: 30, loss: 0.23876327276229858
step: 40, loss: 0.3144698739051819
step: 50, loss: 0.3738466501235962
step: 60, loss: 0.43473950028419495
step: 70, loss: 0.31004709005355835
step: 80, loss: 0.2567903995513916
step: 90, loss: 0.41218918561935425
step: 100, loss: 0.23960426449775696
step: 110, loss: 0.18556863069534302
step: 120, loss: 0.5344347357749939
step: 130, loss: 0.42518842220306396
step: 140, loss: 0.46937116980552673
step: 150, loss: 0.11786145716905594
step: 160, loss: 0.30458864569664
step: 170, loss: 0.16423234343528748
/usr/local/lib/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=0.5932203389830508, f1=0.6186440677966103,
best_f1=0.6186440677966103
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.32910943031311035
```

```
step: 10, loss: 0.08308415859937668
step: 20, loss: 0.21983429789543152
step: 30, loss: 0.18327710032463074
step: 40, loss: 0.18887144327163696
step: 50, loss: 0.17633819580078125
step: 60, loss: 0.07409349828958511
step: 70, loss: 0.23530460894107819
step: 80, loss: 0.21969985961914062
step: 90, loss: 0.21164776384830475
step: 100, loss: 0.1366240233182907
step: 110, loss: 0.3214060068130493
step: 120, loss: 0.13342608511447906
step: 130, loss: 0.0680701956152916
step: 140, loss: 0.21196089684963226
step: 150, loss: 0.11542610824108124
step: 160, loss: 0.15619401633739471
step: 170, loss: 0.10625558346509933
/usr/local/lib/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=0.7087378640776699, f1=0.6808510638297873,
best_f1=0.6808510638297873
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08947759866714478
step: 10, loss: 0.052552517503499985
step: 20, loss: 0.026777267456054688
step: 30, loss: 0.25428125262260437
step: 40, loss: 0.006266262382268906
step: 50, loss: 0.11179666966199875
step: 60, loss: 0.31552764773368835
step: 70, loss: 0.13065794110298157
step: 80, loss: 0.09079151600599289
step: 90, loss: 0.08728902786970139
step: 100, loss: 0.05325012281537056
step: 110, loss: 0.08383198827505112
step: 120, loss: 0.032366059720516205
step: 130, loss: 0.10322288423776627
step: 140, loss: 0.00471894908696413
step: 150, loss: 0.14244243502616882
step: 160, loss: 0.0688997432589531
step: 170, loss: 0.12341451644897461
/usr/local/lib/python3.7/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=0.6929133858267718, f1=0.6894865525672372,
best_f1=0.6808510638297873
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10271412879228592
step: 10, loss: 0.14845699071884155
step: 20, loss: 0.08182254433631897
step: 30, loss: 0.15591581165790558
step: 40, loss: 0.07812397181987762
step: 50, loss: 0.04733269661664963
step: 60, loss: 0.07687314599752426
step: 70, loss: 0.03715270757675171
step: 80, loss: 0.06592180579900742
step: 90, loss: 0.05577554181218147
step: 100, loss: 0.03572552278637886
step: 110, loss: 0.03765907511115074
step: 120, loss: 0.2050389051437378
step: 130, loss: 0.020148856565356255
step: 140, loss: 0.015107783488929272
step: 150, loss: 0.03370283544063568
step: 160, loss: 0.08668313920497894
step: 170, loss: 0.012721177190542221
/usr/local/lib/python3.7/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=0.6981132075471699, f1=0.6789838337182448,
best_f1=0.6808510638297873
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.12502938508987427
step: 10, loss: 0.1005890816450119
step: 20, loss: 0.05679324269294739
step: 30, loss: 0.07273305952548981
step: 40, loss: 0.05319948494434357
step: 50, loss: 0.02277223765850067
step: 60, loss: 0.004695459268987179
step: 70, loss: 0.007990307174623013
step: 80, loss: 0.014384904876351357
step: 90, loss: 0.019949400797486305
step: 100, loss: 0.08481035381555557
step: 110, loss: 0.18740597367286682
step: 120, loss: 0.07998830825090408
step: 130, loss: 0.01500474289059639
step: 140, loss: 0.09017674624919891
```

```
step: 150, loss: 0.018847430124878883
step: 160, loss: 0.0246720053255558
step: 170, loss: 0.04983902350068092
/usr/local/lib/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=0.7117794486215538, f1=0.6733668341708543,
best_f1=0.6733668341708543
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10158319771289825
step: 10, loss: 0.07371331751346588
step: 20, loss: 0.009928129613399506
step: 30, loss: 0.16596955060958862
step: 40, loss: 0.010465229861438274
step: 50, loss: 0.009355448186397552
step: 60, loss: 0.1217384785413742
step: 70, loss: 0.03602183610200882
step: 80, loss: 0.09244001656770706
step: 90, loss: 0.0488775372505188
step: 100, loss: 0.12918338179588318
step: 110, loss: 0.08932068943977356
step: 120, loss: 0.1441441923379898
step: 130, loss: 0.22044584155082703
step: 140, loss: 0.006307033821940422
step: 150, loss: 0.32677334547042847
step: 160, loss: 0.04609477147459984
step: 170, loss: 0.05243730545043945
/usr/local/lib/python3.7/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=0.7047146401985112, f1=0.691542288557214,
best_f1=0.6733668341708543
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0032607796601951122
step: 10, loss: 0.004865690134465694
step: 20, loss: 0.011656852439045906
step: 30, loss: 0.20276589691638947
step: 40, loss: 0.02477818727493286
step: 50, loss: 0.01244254969060421
step: 60, loss: 0.17573432624340057
```

```
step: 70, loss: 0.007588204927742481
step: 80, loss: 0.021175606176257133
step: 90, loss: 0.004527768585830927
step: 100, loss: 0.015345687977969646
step: 110, loss: 0.02097168192267418
step: 120, loss: 0.2229199856519699
step: 130, loss: 0.06539281457662582
step: 140, loss: 0.04239558055996895
step: 150, loss: 0.11759160459041595
step: 160, loss: 0.047328073531389236
step: 170, loss: 0.08342380821704865
/usr/local/lib/python3.7/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=0.6910994764397905, f1=0.6851385390428212,
best_f1=0.6733668341708543
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13040213286876678
step: 10, loss: 0.015657659620046616
step: 20, loss: 0.047851014882326126
step: 30, loss: 0.056282173842191696
step: 40, loss: 0.007741358131170273
step: 50, loss: 0.0012367962626740336
step: 60, loss: 0.006629134528338909
step: 70, loss: 0.006345084868371487
step: 80, loss: 0.0017843652749434114
step: 90, loss: 0.061430543661117554
step: 100, loss: 0.008149714209139347
step: 110, loss: 0.056735921651124954
step: 120, loss: 0.003026908729225397
step: 130, loss: 0.033875517547130585
step: 140, loss: 0.05346100032329559
step: 150, loss: 0.056557513773441315
step: 160, loss: 0.010357336141169071
step: 170, loss: 0.01334922481328249
/usr/local/lib/python3.7/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=0.6947890818858561, f1=0.6893203883495146,
best_f1=0.6733668341708543
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02125287428498268
step: 10, loss: 0.01258022803813219
step: 20, loss: 0.021320069208741188
```

```
step: 30, loss: 0.04191139340400696
step: 40, loss: 0.026741869747638702
step: 50, loss: 0.03320491313934326
step: 60, loss: 0.0034231289755553007
step: 70, loss: 0.144067645072937
step: 80, loss: 0.03252251818776131
step: 90, loss: 0.01982244849205017
step: 100, loss: 0.02632416971027851
step: 110, loss: 0.0013282060390338302
step: 120, loss: 0.03993404656648636
step: 130, loss: 0.02325119450688362
step: 140, loss: 0.0013438964961096644
step: 150, loss: 0.0758049488067627
step: 160, loss: 0.012983150780200958
step: 170, loss: 0.029770130291581154
/usr/local/lib/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=0.712468193384224, f1=0.7007299270072993,
best_f1=0.7007299270072993
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03921610862016678
step: 10, loss: 0.099824458360672
step: 20, loss: 0.006658267229795456
step: 30, loss: 0.0003013374807778746
step: 40, loss: 0.07094080001115799
step: 50, loss: 0.013507931493222713
step: 60, loss: 0.02037145383656025
step: 70, loss: 0.10267829149961472
step: 80, loss: 0.04764719679951668
step: 90, loss: 0.010317937470972538
step: 100, loss: 0.04742836579680443
step: 110, loss: 0.02359568141400814
step: 120, loss: 0.0034139661584049463
step: 130, loss: 0.015640219673514366
step: 140, loss: 0.02050854079425335
step: 150, loss: 0.005713210441172123
step: 160, loss: 0.0004731559893116355
step: 170, loss: 0.006869012955576181
/usr/local/lib/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=0.7139240506329114, f1=0.6889952153110048,
best_f1=0.6889952153110048
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11021200567483902
step: 10, loss: 0.031938180327415466
step: 20, loss: 0.027034347876906395
step: 30, loss: 0.005823586601763964
step: 40, loss: 0.00643521174788475
step: 50, loss: 0.0012281116796657443
step: 60, loss: 0.009435523301362991
step: 70, loss: 0.009424139745533466
step: 80, loss: 0.001834479859098792
step: 90, loss: 0.02726234681904316
step: 100, loss: 0.0015182311180979013
step: 110, loss: 0.007794198580086231
step: 120, loss: 0.12406262755393982
step: 130, loss: 0.012882858514785767
step: 140, loss: 0.07835550606250763
step: 150, loss: 0.002251166384667158
step: 160, loss: 0.0011286924127489328
step: 170, loss: 0.02713657170534134
/usr/local/lib/python3.7/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=0.7061855670103093, f1=0.6930693069306931,
best_f1=0.6889952153110048
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.015827137976884842
step: 10, loss: 0.025357455015182495
step: 20, loss: 0.059215694665908813
step: 30, loss: 0.02081603929400444
step: 40, loss: 0.0007057447801344097
step: 50, loss: 0.0013376092538237572
step: 60, loss: 0.0008455846109427512
step: 70, loss: 0.00821661576628685
step: 80, loss: 0.005156348459422588
step: 90, loss: 0.001961606787517667
step: 100, loss: 0.04657554626464844
step: 110, loss: 0.00020068608864676207
step: 120, loss: 0.0034388869535177946
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 130, loss: 0.001376171363517642
step: 140, loss: 0.001502307364717126
step: 150, loss: 0.007244789972901344
step: 160, loss: 0.022558659315109253
step: 170, loss: 0.008459597826004028
/usr/local/lib/python3.7/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=0.7022900763358779, f1=0.6844660194174758,
best_f1=0.6889952153110048
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00021680226200260222
step: 10, loss: 0.01399984024465084
step: 20, loss: 0.0004843693459406495
step: 30, loss: 0.005038979928940535
step: 40, loss: 0.000654085073620081
step: 50, loss: 0.016326557844877243
step: 60, loss: 0.002814448205754161
step: 70, loss: 0.015187237411737442
step: 80, loss: 0.0038384879007935524
step: 90, loss: 0.004940305836498737
step: 100, loss: 0.0010626019211485982
step: 110, loss: 0.0019251061603426933
step: 120, loss: 0.0012601097114384174
step: 130, loss: 0.003384593641385436
step: 140, loss: 0.0016648524906486273
step: 150, loss: 0.022366929799318314
step: 160, loss: 0.002746609039604664
step: 170, loss: 0.00021098597790114582
/usr/local/lib/python3.7/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=0.7047146401985112, f1=0.6715328467153285,
best_f1=0.6889952153110048
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0017918688245117664
step: 10, loss: 0.0004128095752093941
step: 20, loss: 0.03671392798423767
step: 30, loss: 0.0010758324060589075
step: 40, loss: 0.02126956358551979
step: 50, loss: 0.031761422753334045
step: 60, loss: 0.0009605567320249975
step: 70, loss: 0.0002748368715401739

```
step: 80, loss: 0.0011840835213661194
step: 90, loss: 0.016703037545084953
step: 100, loss: 0.00038011028664186597
step: 110, loss: 0.002530493773519993
step: 120, loss: 0.0019133426249027252
step: 130, loss: 0.002437037881463766
step: 140, loss: 0.0006365231820382178
step: 150, loss: 0.0026805144734680653
step: 160, loss: 0.05644499510526657
step: 170, loss: 0.00046860388829372823
/usr/local/lib/python3.7/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=0.7064935064935066, f1=0.6633165829145728,
best_f1=0.6889952153110048
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0057258908636868
step: 10, loss: 0.15228138864040375
step: 20, loss: 0.0007699186680838466
step: 30, loss: 0.053554050624370575
step: 40, loss: 0.000884641835000366
step: 50, loss: 0.0005486374720931053
step: 60, loss: 0.00017134449444711208
step: 70, loss: 0.0025682293344289064
step: 80, loss: 0.009579821489751339
step: 90, loss: 0.0003564879298210144
step: 100, loss: 0.0015600166516378522
step: 110, loss: 0.015567601658403873
step: 120, loss: 0.016037089750170708
step: 130, loss: 0.0035275090485811234
step: 140, loss: 0.0006470656953752041
step: 150, loss: 0.0007219562539830804
step: 160, loss: 0.005328684579581022
step: 170, loss: 0.0007305228500626981
/usr/local/lib/python3.7/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=0.7079207920792079, f1=0.6903073286052009,
best_f1=0.6889952153110048
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 [00:05, 322.34it/s]
load_f1 = 0.64
real_f1 = 0.631578947368421
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, 250.25it/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, 421kB/s]
Downloading: 100% 232k/232k [00:00<00:00, 1.88MB/s]
Downloading: 100% 268M/268M [00:04<00:00, 64.2MB/s]
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: 0.8175991177558899
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.48768138885498047
step: 20, loss: 0.6124733686447144
step: 30, loss: 0.4956091642379761
step: 40, loss: 0.42658689618110657
step: 50, loss: 0.3137235939502716
step: 60, loss: 0.16500817239284515
step: 70, loss: 0.17832067608833313
step: 80, loss: 0.2352222055196762
step: 90, loss: 0.08111239969730377
step: 100, loss: 0.10318733751773834
step: 110, loss: 0.11954237520694733
step: 120, loss: 0.02995391935110092
step: 130, loss: 0.019679326564073563
step: 140, loss: 0.04131067544221878
step: 150, loss: 0.09811985492706299
step: 160, loss: 0.19351769983768463
step: 170, loss: 0.018493736162781715
step: 180, loss: 0.015678439289331436
step: 190, loss: 0.25382131338119507
step: 200, loss: 0.08592905104160309
step: 210, loss: 0.03986058011651039
step: 220, loss: 0.038219716399908066
step: 230, loss: 0.02574940025806427
/usr/local/lib/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=0.9409141583054627, f1=0.936026936026936,
best_f1=0.936026936026936
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.016407262533903122
step: 10, loss: 0.06462143361568451
step: 20, loss: 0.027987942099571228
step: 30, loss: 0.02713223174214363
step: 40, loss: 0.027321619912981987
step: 50, loss: 0.10899271070957184
step: 60, loss: 0.1627611517906189
step: 70, loss: 0.0719841867685318
step: 80, loss: 0.005912122782319784
step: 90, loss: 0.004701988305896521
step: 100, loss: 0.1593262404203415
step: 110, loss: 0.09518341720104218
step: 120, loss: 0.021126940846443176
step: 130, loss: 0.1845790594816208
step: 140, loss: 0.16231290996074677
step: 150, loss: 0.06919810175895691
step: 160, loss: 0.09169861674308777
step: 170, loss: 0.01438838429749012
step: 180, loss: 0.053592510521411896
step: 190, loss: 0.031611889600753784
step: 200, loss: 0.0977456271648407
step: 210, loss: 0.10297569632530212
step: 220, loss: 0.021211247891187668
step: 230, loss: 0.17935988306999207
/usr/local/lib/python3.7/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=0.9383561643835616, f1=0.9303944315545244,
best_f1=0.936026936026936
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09566619992256165
step: 10, loss: 0.07394496351480484
step: 20, loss: 0.0211337823420763
step: 30, loss: 0.009505582973361015
step: 40, loss: 0.10267505794763565
step: 50, loss: 0.012212100438773632
step: 60, loss: 0.08960198611021042
step: 70, loss: 0.004816174041479826
step: 80, loss: 0.052343737334012985
step: 90, loss: 0.10664352029561996
step: 100, loss: 0.003497040830552578
step: 110, loss: 0.05652996152639389
step: 120, loss: 0.008930608630180359
step: 130, loss: 0.004514060448855162
step: 140, loss: 0.0034262635745108128
```

```
step: 150, loss: 0.016350699588656425
step: 160, loss: 0.0044990424066782
step: 170, loss: 0.027434395626187325
step: 180, loss: 0.0031462342012673616
step: 190, loss: 0.09239613264799118
step: 200, loss: 0.07489701360464096
step: 210, loss: 0.01603688858449459
step: 220, loss: 0.02586829476058483
step: 230, loss: 0.1799786537885666
/usr/local/lib/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=0.951739618406285, f1=0.9527027027027027,
best_f1=0.9527027027027027
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.032123588025569916
step: 10, loss: 0.010619411244988441
step: 20, loss: 0.0017617467092350125
step: 30, loss: 0.0018444781890138984
step: 40, loss: 0.0029032202437520027
step: 50, loss: 0.003927832003682852
step: 60, loss: 0.009855743497610092
step: 70, loss: 0.004376250319182873
step: 80, loss: 0.15019415318965912
step: 90, loss: 0.0072365556843578815
step: 100, loss: 0.019499676302075386
step: 110, loss: 0.022460713982582092
step: 120, loss: 0.01694170944392681
step: 130, loss: 0.001137954881414771
step: 140, loss: 0.008535550907254219
step: 150, loss: 0.001733037643134594
step: 160, loss: 0.05129154399037361
step: 170, loss: 0.005688539706170559
step: 180, loss: 0.09148462116718292
step: 190, loss: 0.006181502714753151
step: 200, loss: 0.011757604777812958
step: 210, loss: 0.0013762564631178975
step: 220, loss: 0.00408053956925869
step: 230, loss: 0.10501404106616974
/usr/local/lib/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=0.9522752497225305, f1=0.9499443826473859,
best_f1=0.9499443826473859
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0017904816195368767
step: 10, loss: 0.002230319892987609
step: 20, loss: 0.002061673905700445
step: 30, loss: 0.001657686079852283
step: 40, loss: 0.0021548226941376925
step: 50, loss: 0.003838352160528302
step: 60, loss: 0.004265397787094116
step: 70, loss: 0.0008768774569034576
step: 80, loss: 0.06611622869968414
step: 90, loss: 0.005629444494843483
step: 100, loss: 0.01813303306698799
step: 110, loss: 0.004162777680903673
step: 120, loss: 0.0443522185087204
step: 130, loss: 0.002482716692611575
step: 140, loss: 0.0005336819449439645
step: 150, loss: 0.00030891611822880805
step: 160, loss: 0.15907453000545502
step: 170, loss: 0.005802148953080177
step: 180, loss: 0.0018558475421741605
step: 190, loss: 0.0007458645850419998
step: 200, loss: 0.002664446597918868
step: 210, loss: 0.05589449405670166
step: 220, loss: 0.002127674175426364
step: 230, loss: 0.0008975974633358419
/usr/local/lib/python3.7/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=0.9513274336283186, f1=0.9465478841870824,
best_f1=0.9499443826473859
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0013320759171620011
step: 10, loss: 0.005977119784802198
step: 20, loss: 0.009712212719023228
step: 30, loss: 0.005017673131078482
step: 40, loss: 0.023931490257382393
step: 50, loss: 0.004293561447411776
step: 60, loss: 0.01120692677795887
```

```
step: 70, loss: 0.0022229244932532231
step: 80, loss: 0.01888044737279415
step: 90, loss: 0.0005508214235305786
step: 100, loss: 0.0012532369000837207
step: 110, loss: 0.10283569991588593
step: 120, loss: 0.004782106727361679
step: 130, loss: 0.0034251767210662365
step: 140, loss: 0.005371833685785532
step: 150, loss: 0.008503125049173832
step: 160, loss: 0.04255789518356323
step: 170, loss: 0.002406417392194271
step: 180, loss: 0.00311990175396204
step: 190, loss: 0.008010409772396088
step: 200, loss: 0.011851289309561253
step: 210, loss: 0.0004351114039309323
step: 220, loss: 0.00024665280943736434
step: 230, loss: 0.0050652832724153996
/usr/local/lib/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=0.9565217391304347, f1=0.9528089887640449,
best_f1=0.9528089887640449
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.015351184643805027
step: 10, loss: 0.0003252899332437664
step: 20, loss: 0.0006937385187484324
step: 30, loss: 0.004750506486743689
step: 40, loss: 0.0019862891640514135
step: 50, loss: 0.10359691828489304
step: 60, loss: 0.0059444899670779705
step: 70, loss: 0.0005841092206537724
step: 80, loss: 0.00048396355123259127
step: 90, loss: 0.0012753031915053725
step: 100, loss: 0.0044107576832175255
step: 110, loss: 0.001322719850577414
step: 120, loss: 0.011274149641394615
step: 130, loss: 0.0014558499678969383
step: 140, loss: 0.0005476147634908557
step: 150, loss: 0.039806898683309555
step: 160, loss: 0.010554872453212738
step: 170, loss: 0.004094318952411413
step: 180, loss: 0.00027335749473422766
```

```
step: 190, loss: 0.0024552468676120043
step: 200, loss: 0.0006688323919661343
step: 210, loss: 0.000194727152120322
step: 220, loss: 0.0016092524165287614
step: 230, loss: 0.0912322998046875
/usr/local/lib/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=0.9605411499436302, f1=0.9561304836895389,
best_f1=0.9561304836895389
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00042689227848313375
step: 10, loss: 0.01841229945421219
step: 20, loss: 0.0007440719055011868
step: 30, loss: 0.0026252414099872112
step: 40, loss: 0.014603800140321255
step: 50, loss: 0.0013074863236397505
step: 60, loss: 0.0010785876074805856
step: 70, loss: 0.0016724420711398125
step: 80, loss: 0.0010958478087559342
step: 90, loss: 0.0004893727600574493
step: 100, loss: 0.0003882997843902558
step: 110, loss: 0.0023681088350713253
step: 120, loss: 0.0009581600315868855
step: 130, loss: 0.06520909816026688
step: 140, loss: 0.004600844345986843
step: 150, loss: 0.007588961161673069
step: 160, loss: 0.010385905392467976
step: 170, loss: 0.04587390646338463
step: 180, loss: 0.005948955193161964
step: 190, loss: 0.010614495724439621
step: 200, loss: 0.016287904232740402
step: 210, loss: 0.0020662047900259495
step: 220, loss: 0.003687122603878379
step: 230, loss: 0.017572369426488876
/usr/local/lib/python3.7/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=0.9550561797752809, f1=0.953229398663697,
best_f1=0.9561304836895389
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006300063687376678
step: 10, loss: 0.0009395845700055361
step: 20, loss: 0.00645094271749258
step: 30, loss: 0.00048344151582568884
step: 40, loss: 0.00028201547684147954
step: 50, loss: 0.0014858637005090714
step: 60, loss: 0.00042895606020465493
step: 70, loss: 0.0005690502002835274
step: 80, loss: 0.1335424929857254
step: 90, loss: 0.004958304110914469
step: 100, loss: 0.0013786726631224155
step: 110, loss: 0.0006006247131153941
step: 120, loss: 0.07394147664308548
step: 130, loss: 0.0011157176923006773
step: 140, loss: 0.11348158866167068
step: 150, loss: 0.00010296410619048402
step: 160, loss: 0.13251543045043945
step: 170, loss: 0.0023446364793926477
step: 180, loss: 0.0004465220554266125
step: 190, loss: 0.0076370276510715485
step: 200, loss: 0.0008541245479136705
step: 210, loss: 0.001851272420026362
step: 220, loss: 0.00378498830832541
step: 230, loss: 0.0015727749560028315
/usr/local/lib/python3.7/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=0.9519553072625698, f1=0.9516310461192351,
best_f1=0.9561304836895389
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009689512080512941
step: 10, loss: 0.00012056235573254526
step: 20, loss: 0.0003643619711510837
step: 30, loss: 0.00027131816023029387
step: 40, loss: 0.0004229997284710407
step: 50, loss: 0.0007016477175056934
step: 60, loss: 0.15766145288944244
step: 70, loss: 0.004148946143686771
step: 80, loss: 0.00016088814300019294
step: 90, loss: 0.00018468669441062957
step: 100, loss: 0.00037572052679024637
step: 110, loss: 0.00010793303226819262
step: 120, loss: 0.006632066331803799
step: 130, loss: 0.0013878496829420328
step: 140, loss: 0.016100499778985977
```

```
step: 150, loss: 0.0008958555990830064
step: 160, loss: 0.0010204797144979239
step: 170, loss: 0.00017730258696246892
step: 180, loss: 0.07792063802480698
step: 190, loss: 0.00044020317727699876
step: 200, loss: 0.0005517056561075151
step: 210, loss: 0.00012402240827213973
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 220, loss: 0.0014148971531540155
step: 230, loss: 0.0011156772961840034
/usr/local/lib/python3.7/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=0.9530201342281879, f1=0.9498327759197325,
best_f1=0.9561304836895389
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.410705438815057e-05
step: 10, loss: 0.00029988630558364093
step: 20, loss: 6.411983486032113e-05
step: 30, loss: 0.00011434495536377653
step: 40, loss: 0.0033872216008603573
step: 50, loss: 0.018929988145828247
step: 60, loss: 0.00013915973249822855
step: 70, loss: 0.00011721943883458152
step: 80, loss: 0.0024579924065619707
step: 90, loss: 0.022534910589456558
step: 100, loss: 0.03387235850095749
step: 110, loss: 0.017602374777197838
step: 120, loss: 0.0012988706585019827
step: 130, loss: 0.0009194189915433526
step: 140, loss: 0.0002273252175655216
step: 150, loss: 6.160608609206975e-05
step: 160, loss: 0.00023119596880860627
step: 170, loss: 0.003795669414103031
step: 180, loss: 0.00016748756752349436
step: 190, loss: 0.0001204651125590317
step: 200, loss: 0.00013690887135453522
step: 210, loss: 6.661348015768453e-05
step: 220, loss: 0.00018720237130764872
step: 230, loss: 0.019067365676164627
/usr/local/lib/python3.7/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=0.9524886877828055, f1=0.9429223744292238,
best_f1=0.9561304836895389
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

```
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0014574566157534719
step: 10, loss: 0.0020920545794069767
step: 20, loss: 0.00010477413889020681
step: 30, loss: 0.00028638640651479363
step: 40, loss: 6.572635902557522e-05
step: 50, loss: 0.00010957277117995545
step: 60, loss: 0.006029557902365923
step: 70, loss: 0.02050534263253212
step: 80, loss: 0.00011974591325270012
step: 90, loss: 0.00022655140492133796
step: 100, loss: 0.00013868998212274164
step: 110, loss: 0.00012773314665537328
step: 120, loss: 0.00040547840762883425
step: 130, loss: 0.0010109430877491832
step: 140, loss: 9.729413432069123e-05
step: 150, loss: 4.538573921308853e-05
step: 160, loss: 0.0021851176861673594
step: 170, loss: 8.974610682344064e-05
step: 180, loss: 0.000428535946412012
step: 190, loss: 0.00016008273814804852
step: 200, loss: 0.00180448975879699
step: 210, loss: 7.816818106221035e-05
step: 220, loss: 0.04670647159218788
step: 230, loss: 0.00011834189353976399
/usr/local/lib/python3.7/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=0.9559322033898305, f1=0.9411764705882352,
best_f1=0.9561304836895389
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.470098302699625e-05
step: 10, loss: 4.948004789184779e-05
step: 20, loss: 0.00046094844583421946
step: 30, loss: 0.04096977785229683
step: 40, loss: 0.0006267678108997643
step: 50, loss: 0.000554982281755656
step: 60, loss: 7.125155389076099e-05
step: 70, loss: 0.00013130488514434546
step: 80, loss: 0.000434188375947997
step: 90, loss: 4.16058610426262e-05
step: 100, loss: 0.000389344641007483
step: 110, loss: 9.791889897314832e-05
step: 120, loss: 0.002712504705414176
step: 130, loss: 0.00027001366834156215
```

```
step: 140, loss: 0.008028123527765274
step: 150, loss: 0.0003508762747514993
step: 160, loss: 0.004020716063678265
step: 170, loss: 5.483698259922676e-05
step: 180, loss: 0.0015379118267446756
step: 190, loss: 4.9974260036833584e-05
step: 200, loss: 7.73123829276301e-05
step: 210, loss: 7.327691855607554e-05
step: 220, loss: 0.00010727307380875573
step: 230, loss: 0.00039043358992785215
/usr/local/lib/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=0.9614512471655329, f1=0.9455782312925171,
best_f1=0.9455782312925171
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.900717223994434e-05
step: 10, loss: 3.617100810515694e-05
step: 20, loss: 0.00169975019525736557
step: 30, loss: 0.00010775875125546008
step: 40, loss: 6.361709529301152e-05
step: 50, loss: 0.000286008056718856
step: 60, loss: 4.177028313279152e-05
step: 70, loss: 5.0594404456205666e-05
step: 80, loss: 0.0007590454188175499
step: 90, loss: 9.073812543647364e-05
step: 100, loss: 0.0013352526584640145
step: 110, loss: 8.710251859156415e-05
step: 120, loss: 3.796309101744555e-05
step: 130, loss: 0.00012207131658215076
step: 140, loss: 0.00012369723117444664
step: 150, loss: 5.47068084415514e-05
step: 160, loss: 5.107554898131639e-05
step: 170, loss: 0.00011324348452035338
step: 180, loss: 0.0004869375843554735
step: 190, loss: 0.0009824811713770032
step: 200, loss: 6.206866237334907e-05
step: 210, loss: 0.0007850779220461845
step: 220, loss: 9.353940549772233e-05
step: 230, loss: 5.649825106956996e-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=0.9614512471655329, f1=0.9478458049886621,
best_f1=0.9455782312925171
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.593407720676623e-05
step: 10, loss: 0.00010115601617144421
step: 20, loss: 0.00011217758583370596
step: 30, loss: 0.0056344326585531235
step: 40, loss: 9.058793511940166e-05
step: 50, loss: 8.516891830367967e-05
step: 60, loss: 3.8205056625884026e-05
step: 70, loss: 0.00017003317771013826
step: 80, loss: 6.594150909222662e-05
step: 90, loss: 4.915034514851868e-05
step: 100, loss: 0.000683110614772886
step: 110, loss: 8.239257294917479e-05
step: 120, loss: 7.443493086611852e-05
step: 130, loss: 0.00011944396828766912
step: 140, loss: 0.0001476663164794445
step: 150, loss: 8.835017069941387e-05
step: 160, loss: 5.2540963224601e-05
step: 170, loss: 5.326302198227495e-05
step: 180, loss: 6.704925908707082e-05
step: 190, loss: 0.0008453863556496799
step: 200, loss: 6.381262210197747e-05
step: 210, loss: 0.007583961356431246
step: 220, loss: 0.0019676911178976297
step: 230, loss: 7.083440141286701e-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=0.9603624009060023, f1=0.9443813847900113,
best_f1=0.9455782312925171
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 [00:08, 277.47it/s]
load_f1 = 0.9614512471655329
real_f1 = 0.9568181818181819
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, 350.46it/s]
```

1.3.2 DBLP-GoogleScholar - Running the matcher

```python
# 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 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: 0.8011091351509094
step: 10, loss: 0.4094763696193695
step: 20, loss: 0.5013610124588013
step: 30, loss: 0.47188159823417664
step: 40, loss: 0.4701383709907532
step: 50, loss: 0.41896742582321167
step: 60, loss: 0.4412967264652252
step: 70, loss: 0.22796793282032013
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 80, loss: 0.38012075424194336
step: 90, loss: 0.37699320912361145
step: 100, loss: 0.25175172090530396
step: 110, loss: 0.11724621057510376
step: 120, loss: 0.18234147131443024
step: 130, loss: 0.053947143256664276
step: 140, loss: 0.32620659470558167
step: 150, loss: 0.05157405883073807
step: 160, loss: 0.15402695536613464
step: 170, loss: 0.31113776564598083
step: 180, loss: 0.2281460016965866
step: 190, loss: 0.06240740790963173
step: 200, loss: 0.25495877861976624
step: 210, loss: 0.10058866441249847
step: 220, loss: 0.05304091051220894
step: 230, loss: 0.12443847954273224
step: 240, loss: 0.08097916096448898
step: 250, loss: 0.07266692072153091
step: 260, loss: 0.02464454062283039
step: 270, loss: 0.026026541367173195
step: 280, loss: 0.22794321179389954
step: 290, loss: 0.1015319973230362
step: 300, loss: 0.15171052515506744
step: 310, loss: 0.05208564177155495
step: 320, loss: 0.1663723886013031
step: 330, loss: 0.1756824254989624
step: 340, loss: 0.26545023918151855
step: 350, loss: 0.09180454909801483
step: 360, loss: 0.09156399965286255
step: 370, loss: 0.1034795492887497
step: 380, loss: 0.22038252651691437
step: 390, loss: 0.04154512658715248
step: 400, loss: 0.01209245529025793
step: 410, loss: 0.11948300898075104
step: 420, loss: 0.07889721542596817
step: 430, loss: 0.1807435303926468
step: 440, loss: 0.21567711234092712
step: 450, loss: 0.09536575525999069
```

```
step: 460, loss: 0.03706255182623863
step: 470, loss: 0.2786632180213928
step: 480, loss: 0.22973865270614624
step: 490, loss: 0.0827040821313858
step: 500, loss: 0.04310613498091698
step: 510, loss: 0.08338332921266556
step: 520, loss: 0.2070103883743286
step: 530, loss: 0.03699749708175659
/usr/local/lib/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=0.9013557737260404, f1=0.8933582787652012,
best_f1=0.8933582787652012
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1574031412601471
step: 10, loss: 0.1608368456363678
step: 20, loss: 0.12236294895410538
step: 30, loss: 0.0907229408621788
step: 40, loss: 0.011439084075391293
step: 50, loss: 0.03750396519899368
step: 60, loss: 0.20558884739875793
step: 70, loss: 0.12601740658283234
step: 80, loss: 0.021048838272690773
step: 90, loss: 0.11371632665395737
step: 100, loss: 0.2932887375354767
step: 110, loss: 0.05956589803099632
step: 120, loss: 0.03438940271735191
step: 130, loss: 0.06715022772550583
step: 140, loss: 0.08333995193243027
step: 150, loss: 0.0675572082400322
step: 160, loss: 0.1126154437661171
step: 170, loss: 0.1944042295217514
step: 180, loss: 0.053837571293115616
step: 190, loss: 0.030809279531240463
step: 200, loss: 0.05626547709107399
step: 210, loss: 0.026686107739806175
step: 220, loss: 0.14528131484985352
step: 230, loss: 0.09338206797838211
step: 240, loss: 0.14466167986392975
step: 250, loss: 0.16870519518852234
step: 260, loss: 0.06865088641643524
step: 270, loss: 0.15890951454639435
```

```
step: 280, loss: 0.10501398146152496
step: 290, loss: 0.04065684601664543
step: 300, loss: 0.02150237187743187
step: 310, loss: 0.2722081243991852
step: 320, loss: 0.10556440055370331
step: 330, loss: 0.05916617065668106
step: 340, loss: 0.011463986709713936
step: 350, loss: 0.160494863986969
step: 360, loss: 0.23519529402256012
step: 370, loss: 0.036400649696588516
step: 380, loss: 0.15533936023712158
step: 390, loss: 0.03985006734728813
step: 400, loss: 0.05404888838529587
step: 410, loss: 0.007148123346269131
step: 420, loss: 0.06604582816362381
step: 430, loss: 0.04226929321885109
step: 440, loss: 0.023208456113934517
step: 450, loss: 0.07595805823802948
step: 460, loss: 0.20758464932441711
step: 470, loss: 0.055867526680231094
step: 480, loss: 0.2919265627861023
step: 490, loss: 0.07113026827573776
step: 500, loss: 0.017365284264087677
step: 510, loss: 0.05286252126097679
step: 520, loss: 0.03595368564128876
step: 530, loss: 0.3289240300655365
/usr/local/lib/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=0.9130434782608695, f1=0.8968140751307656,
best_f1=0.8968140751307656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09123406559228897
step: 10, loss: 0.08336351066827774
step: 20, loss: 0.18406112492084503
step: 30, loss: 0.08104067295789719
step: 40, loss: 0.008738872595131397
step: 50, loss: 0.03694150596857071
step: 60, loss: 0.014366872608661652
step: 70, loss: 0.04913848266005516
step: 80, loss: 0.13380157947540283
step: 90, loss: 0.27382245659828186
```

```
step: 100, loss: 0.04173984378576279
step: 110, loss: 0.050779614597558975
step: 120, loss: 0.028719432651996613
step: 130, loss: 0.06655192375183105
step: 140, loss: 0.09808636456727982
step: 150, loss: 0.07611726969480515
step: 160, loss: 0.0838540643453598
step: 170, loss: 0.018242744728922844
step: 180, loss: 0.09488631039857864
step: 190, loss: 0.04748866707086563
step: 200, loss: 0.014391510747373104
step: 210, loss: 0.21126896142959595
step: 220, loss: 0.012150048278272152
step: 230, loss: 0.09237437695264816
step: 240, loss: 0.05119309946894646
step: 250, loss: 0.1436762809753418
step: 260, loss: 0.021995527669787407
step: 270, loss: 0.01910468190908432
step: 280, loss: 0.05050790682435036
step: 290, loss: 0.06424163281917572
step: 300, loss: 0.06441925466060638
step: 310, loss: 0.16688048839569092
step: 320, loss: 0.13387955725193024
step: 330, loss: 0.041503943502902985
step: 340, loss: 0.006614494137465954
step: 350, loss: 0.08347427099943161
step: 360, loss: 0.004809064324945211
step: 370, loss: 0.029271448031067848
step: 380, loss: 0.027875520288944244
step: 390, loss: 0.018784072250127792
step: 400, loss: 0.07357151061296463
step: 410, loss: 0.12533283233642578
step: 420, loss: 0.10062268376350403
step: 430, loss: 0.11399834603071213
step: 440, loss: 0.05996858701109886
step: 450, loss: 0.19183650612831116
step: 460, loss: 0.1701091080904007
step: 470, loss: 0.021844100207090378
step: 480, loss: 0.027282074093818665
step: 490, loss: 0.003302791155874729
step: 500, loss: 0.12533137202262878
step: 510, loss: 0.0798598900437355
step: 520, loss: 0.012050083838403225
step: 530, loss: 0.04286446049809456
/usr/local/lib/python3.7/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=0.908167974157822, f1=0.9024839006439742,
```

```
best_f1=0.8968140751307656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.027547644451260567
step: 10, loss: 0.02170668914914131
step: 20, loss: 0.044354699552059174
step: 30, loss: 0.1351887732744217
step: 40, loss: 0.01219726912677288
step: 50, loss: 0.038872238248586655
step: 60, loss: 0.029353423044085503
step: 70, loss: 0.013394029811024666
step: 80, loss: 0.017632506787776947
step: 90, loss: 0.053297385573387146
step: 100, loss: 0.04366143420338631
step: 110, loss: 0.0033073769882321358
step: 120, loss: 0.01838819310069084
step: 130, loss: 0.1800723820924759
step: 140, loss: 0.1586807519197464
step: 150, loss: 0.0053738742135465145
step: 160, loss: 0.13066290318965912
step: 170, loss: 0.011658729985356331
step: 180, loss: 0.1187654584646225
step: 190, loss: 0.03327420353889465
step: 200, loss: 0.0013731210492551327
step: 210, loss: 0.0708201602101326
step: 220, loss: 0.008632583543658257
step: 230, loss: 0.25582224130630493
step: 240, loss: 0.012494919821619987
step: 250, loss: 0.009976860135793686
step: 260, loss: 0.03916347399353981
step: 270, loss: 0.023976163938641548
step: 280, loss: 0.00785074196755886
step: 290, loss: 0.035153038799762726
step: 300, loss: 0.053665533661842346
step: 310, loss: 0.015768423676490784
step: 320, loss: 0.01986530050635338
step: 330, loss: 0.036623015999794006
step: 340, loss: 0.013829429633915424
step: 350, loss: 0.003939576912671328
step: 360, loss: 0.018206119537353516
step: 370, loss: 0.00861439760774374
step: 380, loss: 0.014237086288630962
step: 390, loss: 0.029704904183745384
step: 400, loss: 0.057538941502571106
step: 410, loss: 0.06420572102069855
step: 420, loss: 0.033313170075416565
step: 430, loss: 0.023331306874752045
```

```
step: 440, loss: 0.0473497249186039
step: 450, loss: 0.010942271910607815
step: 460, loss: 0.004079486709088087
step: 470, loss: 0.001995975384488702
step: 480, loss: 0.09905350208282471
step: 490, loss: 0.07757411897182465
step: 500, loss: 0.01892431639134884
step: 510, loss: 0.13028670847415924
step: 520, loss: 0.11982928961515427
step: 530, loss: 0.0035184286534786224
/usr/local/lib/python3.7/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=0.9127769919849128, f1=0.9049858889934148,
best_f1=0.8968140751307656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005767287686467171
step: 10, loss: 0.025439182296395302
step: 20, loss: 0.033258967101573944
step: 30, loss: 0.00994283240288496
step: 40, loss: 0.01297435536980629
step: 50, loss: 0.08490829169750214
step: 60, loss: 0.025212805718183517
step: 70, loss: 0.0029342707712203264
step: 80, loss: 0.0032709757797420025
step: 90, loss: 0.09686999022960663
step: 100, loss: 0.006791945081204176
step: 110, loss: 0.0026916726492345333
step: 120, loss: 0.039199814200401306
step: 130, loss: 0.028729882091283798
step: 140, loss: 0.010595960542559624
step: 150, loss: 0.006528983823955059
step: 160, loss: 0.20252206921577454
step: 170, loss: 0.05761932581663132
step: 180, loss: 0.10069642961025238
step: 190, loss: 0.0019680068362504244
step: 200, loss: 0.0014174594543874264
step: 210, loss: 0.0017507896991446614
step: 220, loss: 0.009136438369750977
step: 230, loss: 0.0006471445667557418
step: 240, loss: 0.0022335408721119165
step: 250, loss: 0.023091452196240425
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 260, loss: 0.030401088297367096
step: 270, loss: 0.0626947358250618
step: 280, loss: 0.031107190996408463
```

```
step: 290, loss: 0.006311541423201561
step: 300, loss: 0.005753392353653908
step: 310, loss: 0.000747904006857425
step: 320, loss: 0.11741246283054352
step: 330, loss: 0.025279046967625618
step: 340, loss: 0.0024649559054523706
step: 350, loss: 0.011149241589009762
step: 360, loss: 0.003342448966577649
step: 370, loss: 0.003221308346837759
step: 380, loss: 0.017818614840507507
step: 390, loss: 0.07687628269195557
step: 400, loss: 0.0010547630954533815
step: 410, loss: 0.002301793312653899
step: 420, loss: 0.00835672952234745
step: 430, loss: 0.027272727340459824
step: 440, loss: 0.056749127805233
step: 450, loss: 0.053509876132011414
step: 460, loss: 0.023668775334954262
step: 470, loss: 0.005449789110571146
step: 480, loss: 0.01300622895359993
step: 490, loss: 0.029811495915055275
step: 500, loss: 0.015401945449411869
step: 510, loss: 0.02126743085682392
step: 520, loss: 0.13668496906757355
step: 530, loss: 0.04784318804740906
/usr/local/lib/python3.7/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=0.9047844623401232, f1=0.8948356807511737,
best_f1=0.8968140751307656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.011453904211521149
step: 10, loss: 0.0531328022480011
step: 20, loss: 0.0008801231742836535
step: 30, loss: 0.008659749291837215
step: 40, loss: 0.004796651192009449
step: 50, loss: 0.0435660220682621
step: 60, loss: 0.08512244373559952
step: 70, loss: 0.013293547555804253
step: 80, loss: 0.0009493406396359205
step: 90, loss: 0.10472079366445541
step: 100, loss: 0.041333701461553574
step: 110, loss: 0.0023544265422970057
step: 120, loss: 0.24108120799064636
step: 130, loss: 0.010298161767423153
step: 140, loss: 0.04167357459664345
```

```
step: 150, loss: 0.0030847815796732903
step: 160, loss: 0.0041227880865335464
step: 170, loss: 0.029088007286190987
step: 180, loss: 0.0006236339686438441
step: 190, loss: 0.05562610924243927
step: 200, loss: 0.0004080980725120753
step: 210, loss: 0.05606839433312416
step: 220, loss: 0.0033401651307940483
step: 230, loss: 0.000655224546790123
step: 240, loss: 0.07311278581619263
step: 250, loss: 0.013152903877198696
step: 260, loss: 0.0030132823158055544
step: 270, loss: 0.00490047549828887
step: 280, loss: 0.003141881665214896
step: 290, loss: 0.004030011128634214
step: 300, loss: 0.0005602266173809767
step: 310, loss: 0.0020160714630037546
step: 320, loss: 0.005374527536332607
step: 330, loss: 0.053713493049144745
step: 340, loss: 0.012483927421271801
step: 350, loss: 0.018159061670303345
step: 360, loss: 0.002042977139353752
step: 370, loss: 0.018476437777280807
step: 380, loss: 0.004745409358292818
step: 390, loss: 0.007156719453632832
step: 400, loss: 0.0035260627046227455
step: 410, loss: 0.00035289840889163315
step: 420, loss: 0.013423705473542213
step: 430, loss: 0.0009401264833286405
step: 440, loss: 0.01749960333108902
step: 450, loss: 0.015002300031483173
step: 460, loss: 0.18656183779239655
step: 470, loss: 0.004521525930613279
step: 480, loss: 0.0021079860161989927
step: 490, loss: 0.004994198214262724
step: 500, loss: 0.07429884374141693
step: 510, loss: 0.06988125294446945
step: 520, loss: 0.002113231224939227
step: 530, loss: 0.005268416367471218
/usr/local/lib/python3.7/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=0.9043397106859542, f1=0.8936567164179104,
best_f1=0.8968140751307656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10020275413990021
```

```
step: 10, loss: 0.03665263578295708
step: 20, loss: 0.14735007286071777
step: 30, loss: 0.006900786887854338
step: 40, loss: 0.0008279496105387807
step: 50, loss: 0.008198057301342487
step: 60, loss: 0.029353152960538864
step: 70, loss: 0.0028449606616050005
step: 80, loss: 0.0015807339223101735
step: 90, loss: 0.006028399337083101
step: 100, loss: 0.0036473332438617945
step: 110, loss: 0.0012323021655902267
step: 120, loss: 0.0002460202085785568
step: 130, loss: 0.01427033357322216
step: 140, loss: 0.002064867177978158
step: 150, loss: 0.06445791572332382
step: 160, loss: 0.007291924208402634
step: 170, loss: 0.003135361708700657
step: 180, loss: 0.052246659994125366
step: 190, loss: 0.0037220637314021587
step: 200, loss: 0.06364136189222336
step: 210, loss: 0.00562805263325572
step: 220, loss: 0.0007437487365677953
step: 230, loss: 0.0013711616629734635
step: 240, loss: 0.0251578688621521
step: 250, loss: 0.035305991768836975
step: 260, loss: 0.0034232083708047867
step: 270, loss: 0.00047837733291089535
step: 280, loss: 0.06139974296092987
step: 290, loss: 0.0024832452181726694
step: 300, loss: 0.014605535194277763
step: 310, loss: 0.0003253151080571115
step: 320, loss: 0.012824947014451027
step: 330, loss: 0.05135160684585571
step: 340, loss: 0.008384454064071178
step: 350, loss: 0.0031315318774431944
step: 360, loss: 0.0015196213498711586
step: 370, loss: 0.01813136227428913
step: 380, loss: 0.004227330908179283
step: 390, loss: 0.008103678934276104
step: 400, loss: 0.0005424093687906861
step: 410, loss: 0.0003245080588385463
step: 420, loss: 0.08139490336179733
step: 430, loss: 0.0003492562682367861
step: 440, loss: 0.003178022801876068
step: 450, loss: 0.0018185763619840145
step: 460, loss: 0.0026343914214521646
step: 470, loss: 0.01053392980247736
step: 480, loss: 0.003625868121162057
```

```
step: 490, loss: 0.007442348171025515
step: 500, loss: 0.00013113350723870099
step: 510, loss: 0.05032581090927124
step: 520, loss: 0.001293982146307826
step: 530, loss: 0.01487792283296585
/usr/local/lib/python3.7/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=0.906791569086651, f1=0.9037106622827618,
best_f1=0.8968140751307656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004483393859118223
step: 10, loss: 0.07473433762788773
step: 20, loss: 0.00545307993888855
step: 30, loss: 0.05584683641791344
step: 40, loss: 0.005192270968109369
step: 50, loss: 0.0027667430695146322
step: 60, loss: 0.0921572893857956
step: 70, loss: 0.07533104717731476
step: 80, loss: 0.007490553427487612
step: 90, loss: 0.001105765113607049
step: 100, loss: 0.0039633349515497684
step: 110, loss: 0.0011004726402461529
step: 120, loss: 0.0013790512457489967
step: 130, loss: 0.017387645319104195
step: 140, loss: 0.00010512234439374879
step: 150, loss: 0.0025507682003080845
step: 160, loss: 0.005220004823058844
step: 170, loss: 0.0061845239251852036
step: 180, loss: 0.0010710828937590122
step: 190, loss: 0.055417753756046295
step: 200, loss: 0.005090763326734304
step: 210, loss: 0.09503105282783508
step: 220, loss: 0.002585142385214567
step: 230, loss: 0.0033775942865759134
step: 240, loss: 0.05374512821435928
step: 250, loss: 0.002030494622886181
step: 260, loss: 0.09171643853187561
step: 270, loss: 0.00017973432841245085
step: 280, loss: 0.000791457830928266
step: 290, loss: 0.012476027011871338
step: 300, loss: 0.00028399238362908363
step: 310, loss: 0.039360105991363525
step: 320, loss: 0.0005776117905043066
step: 330, loss: 0.0010263252770528197
step: 340, loss: 0.0006994931027293205
```

```
step: 350, loss: 0.0011453621555119753
step: 360, loss: 0.006108297035098076
step: 370, loss: 0.0018567141378298402
step: 380, loss: 0.08621316403150558
step: 390, loss: 0.005380840972065926
step: 400, loss: 0.03342701122164726
step: 410, loss: 0.06130418926477432
step: 420, loss: 0.03461742028594017
step: 430, loss: 0.0035952709149569273
step: 440, loss: 0.0016601107781752944
step: 450, loss: 0.000713265675585717
step: 460, loss: 0.010914869606494904
step: 470, loss: 0.0008784868405200541
step: 480, loss: 0.00696626678109169
step: 490, loss: 0.0006881196168251336
step: 500, loss: 0.0002753747976385057
step: 510, loss: 0.003028387436643243
step: 520, loss: 0.003821052610874176
step: 530, loss: 0.0002987275074701756
/usr/local/lib/python3.7/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=0.9056956115779645, f1=0.8996733551096594,
best_f1=0.8968140751307656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007860878249630332
step: 10, loss: 0.000794224557466805
step: 20, loss: 0.01106479112058878
step: 30, loss: 0.016319675371050835
step: 40, loss: 0.37972190976142883
step: 50, loss: 0.0028631589375436306
step: 60, loss: 0.001293072127737105
step: 70, loss: 0.009812279604375362
step: 80, loss: 0.003531040856614709
step: 90, loss: 0.05121295154094696
step: 100, loss: 0.001568412990309298
step: 110, loss: 0.00482749380171299
step: 120, loss: 0.0026272046379745007
step: 130, loss: 0.0007150016026571393
step: 140, loss: 0.0034650657325983047
step: 150, loss: 0.0008481511613354087
step: 160, loss: 0.001243661274202168
step: 170, loss: 0.0009124952484853566
step: 180, loss: 0.00017059850506484509
step: 190, loss: 0.0106691038236022
step: 200, loss: 0.0019172753673046827
```

```
step: 210, loss: 0.00011217398423468694
step: 220, loss: 0.0002756469475571066
step: 230, loss: 0.0008619948639534414
step: 240, loss: 0.006322868634015322
step: 250, loss: 0.003425262402743101
step: 260, loss: 0.0012219962663948536
step: 270, loss: 0.00669822096824646
step: 280, loss: 8.328596595674753e-05
step: 290, loss: 0.00011955325317103416
step: 300, loss: 0.00183952902443707
step: 310, loss: 0.0005118089029565454
step: 320, loss: 0.0011141003342345357
step: 330, loss: 0.0010429049143567681
step: 340, loss: 0.002311899559572339
step: 350, loss: 0.0006570058176293969
step: 360, loss: 0.01171304751187563
step: 370, loss: 0.0031731559429317713
step: 380, loss: 0.00033224382787011564
step: 390, loss: 0.0006474981200881302
step: 400, loss: 0.0034913201816380024
step: 410, loss: 0.2726477384567261
step: 420, loss: 0.015048308297991753
step: 430, loss: 0.002175979781895876
step: 440, loss: 0.003519258927553892
step: 450, loss: 0.00026960301329381764
step: 460, loss: 0.05244944244623184
step: 470, loss: 0.0005386017146520317
step: 480, loss: 0.0006639832863584161
step: 490, loss: 0.00012737048382405192
step: 500, loss: 0.00043377792462706566
step: 510, loss: 0.025932354852557182
step: 520, loss: 0.00024702560040168464
step: 530, loss: 0.009181316941976547
/usr/local/lib/python3.7/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=0.9079008882655447, f1=0.8989710009354537,
best_f1=0.8968140751307656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012027104385197163
step: 10, loss: 0.00044475888716988266
step: 20, loss: 0.00025433904374949634
step: 30, loss: 0.00010400939936516806
step: 40, loss: 0.04024670645594597
step: 50, loss: 0.001650461577810347
step: 60, loss: 0.0003563447098713368
```

```
step: 70, loss: 0.012703871354460716
step: 80, loss: 0.0006266163545660675
step: 90, loss: 0.0070595950819551945
step: 100, loss: 0.0012273190077394247
step: 110, loss: 0.0013886252418160439
step: 120, loss: 0.00013402875629253685
step: 130, loss: 0.00021823772112838924
step: 140, loss: 0.00046664304682053626
step: 150, loss: 0.0009944570483639836
step: 160, loss: 9.066024358617142e-05
step: 170, loss: 0.0007456735474988818
step: 180, loss: 0.03243149444460869
step: 190, loss: 0.00017155332898255438
step: 200, loss: 0.0024174784775823355
step: 210, loss: 0.0009720802772790194
step: 220, loss: 0.00030745219555683434
step: 230, loss: 0.00047114960034377873
step: 240, loss: 9.287094871979207e-05
step: 250, loss: 0.01990610547363758
step: 260, loss: 0.0010035809827968478
step: 270, loss: 0.03073889948427677
step: 280, loss: 0.000982161145657301
step: 290, loss: 0.010238186456263065
step: 300, loss: 0.0036721983924508095
step: 310, loss: 0.00114213966298848
step: 320, loss: 0.0008982871077023447
step: 330, loss: 0.006809975486248732
step: 340, loss: 0.0033795060589909554
step: 350, loss: 0.0018241096986457705
step: 360, loss: 0.0004676792596001178
step: 370, loss: 0.003342282259836793
step: 380, loss: 0.000341543200192973
step: 390, loss: 0.00031583139207214117
step: 400, loss: 0.0012647591065615416
step: 410, loss: 0.00035454853787086904
step: 420, loss: 0.009085875004529953
step: 430, loss: 0.000260374799836427
step: 440, loss: 0.00016762290033511817
step: 450, loss: 0.0011714767897501588
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 460, loss: 0.007588534615933895
step: 470, loss: 0.00012571913248393685
step: 480, loss: 0.0002861169050447643
step: 490, loss: 0.10897193104028702
step: 500, loss: 0.01295152772217989
step: 510, loss: 0.0014518669340759516
step: 520, loss: 0.0016862699994817376
step: 530, loss: 5.083536962047219e-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=0.9118738404452691, f1=0.9005090236001851,
best_f1=0.8968140751307656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002138124778866768
step: 10, loss: 0.0006059369770810008
step: 20, loss: 0.0006647186237387359
step: 30, loss: 0.014037054032087326
step: 40, loss: 0.00010604114504531026
step: 50, loss: 0.01801149547100067
step: 60, loss: 0.00028136346372775733
step: 70, loss: 6.748777377652004e-05
step: 80, loss: 6.999884499236941e-05
step: 90, loss: 0.00046857865527272224
step: 100, loss: 0.0003007351770065725
step: 110, loss: 0.00016637569933664054
step: 120, loss: 0.004598225001245737
step: 130, loss: 0.0003116761799901724
step: 140, loss: 0.0024157315492630005
step: 150, loss: 0.0006156233721412718
step: 160, loss: 0.0010915981838479638
step: 170, loss: 0.001724525704048574
step: 180, loss: 0.0011091240448877215
step: 190, loss: 0.001470607123337686
step: 200, loss: 0.07523947209119797
step: 210, loss: 0.00015155003347899765
step: 220, loss: 0.0013172010658308864
step: 230, loss: 0.003426524344831705
step: 240, loss: 0.00010654663492459804
step: 250, loss: 0.0013769888319075108
step: 260, loss: 7.700416608713567e-05
step: 270, loss: 0.005159109365195036
step: 280, loss: 0.0019488735124468803
step: 290, loss: 0.0009749167365953326
step: 300, loss: 0.024365298449993134
step: 310, loss: 0.006453773472458124
step: 320, loss: 0.05740124732255936
step: 330, loss: 0.0009971236577257514
step: 340, loss: 0.005010729655623436
step: 350, loss: 0.2239256650209427
step: 360, loss: 0.0008896042127162218
step: 370, loss: 0.0016423359047621489
step: 380, loss: 0.00030625335057266057
step: 390, loss: 0.13502518832683563
```

```
step: 400, loss: 0.00013957357441540807
step: 410, loss: 0.004173182416707277
step: 420, loss: 0.0022506373934447765
step: 430, loss: 0.02569248527288437
step: 440, loss: 0.00047407718375325203
step: 450, loss: 0.00017564267909619957
step: 460, loss: 0.0009787128074094653
step: 470, loss: 0.0013726719189435244
step: 480, loss: 0.0009199604392051697
step: 490, loss: 0.005976362619549036
step: 500, loss: 0.0002161307929782197
step: 510, loss: 0.0002449729945510626
step: 520, loss: 7.174871279858053e-05
step: 530, loss: 0.00015206815442070365
/usr/local/lib/python3.7/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=0.9017074296262113, f1=0.8963922294172062,
best_f1=0.8968140751307656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000462401716504246
step: 10, loss: 5.76851271034684e-05
step: 20, loss: 5.092595165478997e-05
step: 30, loss: 0.001057518646121025
step: 40, loss: 0.014859489165246487
step: 50, loss: 0.004226174205541611
step: 60, loss: 0.0008980490383692086
step: 70, loss: 0.006500354968011379
step: 80, loss: 0.0011043711565434933
step: 90, loss: 0.0011087014572694898
step: 100, loss: 0.0004045462410431355
step: 110, loss: 0.0006413396331481636
step: 120, loss: 0.00029710761737707843
step: 130, loss: 0.0006593340076506138
step: 140, loss: 0.0015317685902118683
step: 150, loss: 0.0004147816507611424
step: 160, loss: 0.00039257106254808605
step: 170, loss: 0.00016289706400129944
step: 180, loss: 0.0005404561525210738
step: 190, loss: 0.0001479985221521929
step: 200, loss: 0.006934932433068752
step: 210, loss: 0.0008812513551674783
step: 220, loss: 0.00016721953579690307
step: 230, loss: 0.0007649441249668598
step: 240, loss: 0.001641020062379539
step: 250, loss: 0.005277357995510101
```

```
step: 260, loss: 0.00043320871191099286
step: 270, loss: 0.021292220801115036
step: 280, loss: 0.08069636672735214
step: 290, loss: 0.001815298805013299
step: 300, loss: 0.00034012910327874124
step: 310, loss: 0.0013114243047311902
step: 320, loss: 0.0007006182568147779
step: 330, loss: 7.009994442341849e-05
step: 340, loss: 0.004431867506355047
step: 350, loss: 0.0012508811196312308
step: 360, loss: 0.0022211940959095955
step: 370, loss: 0.00028842707979492843
step: 380, loss: 0.005111939739435911
step: 390, loss: 8.855982014210895e-05
step: 400, loss: 0.00036457227542996407
step: 410, loss: 0.0008507744641974568
step: 420, loss: 0.000889709684997797
step: 430, loss: 0.0004576884675770998
step: 440, loss: 0.0018056154949590564
step: 450, loss: 0.0005497379461303353
step: 460, loss: 0.09548019617795944
step: 470, loss: 7.929639832582325e-05
step: 480, loss: 0.0005942209390923381
step: 490, loss: 0.017584212124347687
step: 500, loss: 0.0006851693615317345
step: 510, loss: 0.0008564126910641789
step: 520, loss: 0.007644485682249069
step: 530, loss: 0.0007515870966017246
/usr/local/lib/python3.7/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=0.912901723334886, f1=0.9002320185614848,
best_f1=0.8968140751307656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00023486942518502474
step: 10, loss: 0.00015160605835262686
step: 20, loss: 0.0014265564968809485
step: 30, loss: 0.027262559160590172
step: 40, loss: 0.0001555993512738496
step: 50, loss: 0.0003468201030045748
step: 60, loss: 0.0008500547846779227
step: 70, loss: 0.00039402872789651155
step: 80, loss: 0.008635240606963634
step: 90, loss: 0.00012718835205305368
step: 100, loss: 0.00011351380089763552
step: 110, loss: 0.0012660089414566755
```

```
step: 120, loss: 0.00233538169413805
step: 130, loss: 0.00032158184330910444
step: 140, loss: 0.00021973997354507446
step: 150, loss: 0.0003119955654256046
step: 160, loss: 0.0001303623866988346
step: 170, loss: 0.00017776225286070257
step: 180, loss: 0.0001828503591241315
step: 190, loss: 0.0007190638571046293
step: 200, loss: 0.004185650497674942
step: 210, loss: 0.0007129572331905365
step: 220, loss: 0.04137511923909187
step: 230, loss: 0.0014376882463693619
step: 240, loss: 7.464802911272272e-05
step: 250, loss: 0.02109583094716072
step: 260, loss: 9.772894554771483e-05
step: 270, loss: 0.0005836858763359487
step: 280, loss: 0.0031298946123570204
step: 290, loss: 0.0007704122108407319
step: 300, loss: 0.001720533473417163
step: 310, loss: 0.00029982696287333965
step: 320, loss: 0.0009812124771997333
step: 330, loss: 0.0005639774026349187
step: 340, loss: 0.00024073607346508652
step: 350, loss: 0.04685630276799202
step: 360, loss: 7.274426752701402e-05
step: 370, loss: 0.019682148471474648
step: 380, loss: 0.00013376415881793946
step: 390, loss: 0.00015560285828541964
step: 400, loss: 0.0003104773932136595
step: 410, loss: 9.43216509767808e-05
step: 420, loss: 0.00018888567865360528
step: 430, loss: 0.00036496366637419462
step: 440, loss: 0.0005284960498102009
step: 450, loss: 7.432911661453545e-05
step: 460, loss: 3.725516580743715e-05
step: 470, loss: 0.0002152867818949744
step: 480, loss: 0.0010838292073458433
step: 490, loss: 0.017170101404190063
step: 500, loss: 0.0005109385820105672
step: 510, loss: 7.287855260074139e-05
step: 520, loss: 0.00042660036706365645
step: 530, loss: 0.0003526130458340049
/usr/local/lib/python3.7/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=0.9104477611940298, f1=0.8992537313432837,
best_f1=0.8968140751307656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

```
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00014881082461215556
step: 10, loss: 0.0019480952760204673
step: 20, loss: 0.0006239436916075647
step: 30, loss: 0.00019384131883271039
step: 40, loss: 0.0014435012126341462
step: 50, loss: 0.001404112670570612
step: 60, loss: 0.0003058084403164685
step: 70, loss: 5.983271694276482e-05
step: 80, loss: 0.00013337926066014916
step: 90, loss: 4.4034306483808905e-05
step: 100, loss: 0.00012727915600407869
step: 110, loss: 0.00016766710905358195
step: 120, loss: 6.945379573153332e-05
step: 130, loss: 0.00017696904251351953
step: 140, loss: 0.011535638011991978
step: 150, loss: 0.0002569708158262074
step: 160, loss: 0.0008036610088311136
step: 170, loss: 0.021654469892382622
step: 180, loss: 7.939052011352032e-05
step: 190, loss: 0.0015738976653665304
step: 200, loss: 5.4934280342422426e-05
step: 210, loss: 0.0010435881558805704
step: 220, loss: 0.0026950729079544544
step: 230, loss: 0.0001557792566018179
step: 240, loss: 0.0006527315126731992
step: 250, loss: 5.128556949784979e-05
step: 260, loss: 5.976195097900927e-05
step: 270, loss: 0.0008545691380277276
step: 280, loss: 0.0003716648498084396
step: 290, loss: 0.09878689795732498
step: 300, loss: 0.0003764159919228405
step: 310, loss: 0.0006418786360882223
step: 320, loss: 0.0008460166282020509
step: 330, loss: 0.00035532668698579073
step: 340, loss: 8.888105367077515e-05
step: 350, loss: 0.001974627608433366
step: 360, loss: 0.0012330625904724002
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 370, loss: 8.421078382525593e-05
step: 380, loss: 5.312703797244467e-05
step: 390, loss: 0.0013907612301409245
step: 400, loss: 0.0002624897169880569
step: 410, loss: 0.00018946525233332068
step: 420, loss: 0.0012746648862957954
step: 430, loss: 0.0021619643084704876
step: 440, loss: 8.071947377175093e-05
```

```
step: 450, loss: 9.342267003376037e-05
step: 460, loss: 0.00017856994236353785
step: 470, loss: 0.002208513906225562
step: 480, loss: 7.309616194106638e-05
step: 490, loss: 0.0001725249458104372
step: 500, loss: 0.002044475171715021
step: 510, loss: 0.0002834609185811132
step: 520, loss: 6.167140236357227e-05
step: 530, loss: 0.0008737337775528431
/usr/local/lib/python3.7/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=0.9107142857142857, f1=0.8989186647860837,
best_f1=0.8968140751307656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000190291742910631
step: 10, loss: 0.00017111128545366228
step: 20, loss: 0.00036803469993174076
step: 30, loss: 0.0001984510599868372
step: 40, loss: 0.0001813837152440101
step: 50, loss: 0.00012913835234940052
step: 60, loss: 5.824149411637336e-05
step: 70, loss: 0.00035351610858924687
step: 80, loss: 0.00014592535444535315
step: 90, loss: 9.114453860092908e-05
step: 100, loss: 0.00877261534333229
step: 110, loss: 0.008980238810181618
step: 120, loss: 0.0035568701568990946
step: 130, loss: 0.0031749403569847345
step: 140, loss: 8.835265907691792e-05
step: 150, loss: 0.00016859632160048932
step: 160, loss: 0.00031747852335684
step: 170, loss: 0.0007335413829423487
step: 180, loss: 0.01063623744994402
step: 190, loss: 0.000723932171240449
step: 200, loss: 0.0010178685188293457
step: 210, loss: 0.0045797266606607437
step: 220, loss: 0.0001369740639347583
step: 230, loss: 0.00037233642069622874
step: 240, loss: 0.0003670285805128515
step: 250, loss: 0.00011135335080325603
step: 260, loss: 0.0008060384425334632
step: 270, loss: 0.0004333837714511901
step: 280, loss: 0.0002685687504708767
step: 290, loss: 0.0009998823516070843
step: 300, loss: 0.002437267452478409
```

```
step: 310, loss: 0.00020827670232392848
step: 320, loss: 0.000488179997773841
step: 330, loss: 0.00898405909538269
step: 340, loss: 0.00037790031637996435
step: 350, loss: 0.00034688450978137553
step: 360, loss: 0.0019810895901173353
step: 370, loss: 0.005147247575223446
step: 380, loss: 0.00013838287850376219
step: 390, loss: 8.173519745469093e-05
step: 400, loss: 0.0007629881147295237
step: 410, loss: 0.0002718088508117944
step: 420, loss: 0.000141739787068218
step: 430, loss: 0.00038167755701579154
step: 440, loss: 0.00327146309427917
step: 450, loss: 0.00814011599868536
step: 460, loss: 0.0007116387714631855
step: 470, loss: 9.660298383096233e-05
step: 480, loss: 0.0005006479914300144
step: 490, loss: 0.00013555549958255142
step: 500, loss: 0.006457564886659384
step: 510, loss: 0.00019489377154968668
step: 520, loss: 0.0006021123263053596
step: 530, loss: 8.834028267301619e-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=0.9100329722091381, f1=0.8991517436380774,
best_f1=0.8968140751307656
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.")
5742it [00:17, 335.68it/s]
load_f1 = 0.9079754601226995
real_f1 = 0.908235294117647
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, 382.96it/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 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: 0.8810604214668274
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=0.30769230769230765, f1=0.37209302325581395,
best_f1=0.37209302325581395
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.37269771099090576
/usr/local/lib/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=0.4150943396226415, f1=0.3389830508474576,
best_f1=0.3389830508474576
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2817763090133667
/usr/local/lib/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=0.4666666666666667, f1=0.41935483870967744,
best_f1=0.41935483870967744
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3187721371650696
/usr/local/lib/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=0.49122807017543857, f1=0.41935483870967744,
best_f1=0.41935483870967744
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.17136512696743011
/usr/local/lib/python3.7/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=0.4827586206896552, f1=0.41935483870967744,
best_f1=0.41935483870967744
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.25021272897720337
/usr/local/lib/python3.7/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=0.4745762711864407, f1=0.3870967741935483,
best_f1=0.41935483870967744
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1913338452577591
/usr/local/lib/python3.7/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=0.4897959183673469, f1=0.42857142857142855,
best_f1=0.41935483870967744
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.324727863073349
/usr/local/lib/python3.7/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=0.4864864864864865, f1=0.43902439024390244,
best_f1=0.41935483870967744
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.16103851795196533
/usr/local/lib/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=0.5, f1=0.4680851063829786, best_f1=0.4680851063829786
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.24904607236385345
/usr/local/lib/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=0.5128205128205129, f1=0.4680851063829786,
best_f1=0.4680851063829786
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.18968364596366882
/usr/local/lib/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=0.5142857142857143, f1=0.43902439024390244,
best_f1=0.43902439024390244
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.32987794280052185
/usr/local/lib/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=0.5294117647058824, f1=0.4736842105263159,
best_f1=0.4736842105263159
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.18176516890525818
/usr/local/lib/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=0.5652173913043478, f1=0.43999999999999995,
best_f1=0.43999999999999995
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2146664410829544
/usr/local/lib/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=0.5777777777777778, f1=0.43999999999999995,
best_f1=0.43999999999999995
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.17742882668972015
/usr/local/lib/python3.7/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=0.5777777777777778, f1=0.43999999999999995,
best_f1=0.43999999999999995
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, 136656.52it/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 = 0.5365853658536585
real_f1 = 0.5777777777777778
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 [00:16, 266.61it/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 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: 0.8095021843910217
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.49099940061569214
step: 20, loss: 0.5960636138916016
step: 30, loss: 0.5064984560012817
step: 40, loss: 0.4307613670825958
step: 50, loss: 0.3751507103443146
step: 60, loss: 0.2683727741241455
step: 70, loss: 0.39262694120407104
step: 80, loss: 0.1327124685049057
step: 90, loss: 0.011256000027060509
step: 100, loss: 0.26104435324668884
step: 110, loss: 0.04822683334350586
step: 120, loss: 0.053421854972839355

141
```

```
step: 130, loss: 0.019489198923110962
step: 140, loss: 0.09274432808160782
step: 150, loss: 0.04538514465093613
step: 160, loss: 0.1455625742673874
step: 170, loss: 0.010842147283256054
step: 180, loss: 0.010646922513842583
step: 190, loss: 0.1437896341085434
step: 200, loss: 0.05546094849705696
step: 210, loss: 0.029780222102999687
step: 220, loss: 0.01745501533150673
step: 230, loss: 0.008367025293409824
/usr/local/lib/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=0.9557321225879684, f1=0.9554285714285714,
best_f1=0.9554285714285714
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004367798566818237
step: 10, loss: 0.03353540971875191
step: 20, loss: 0.004971010610461235
step: 30, loss: 0.0011425254633650184
step: 40, loss: 0.005080432631075382
step: 50, loss: 0.0020157203543931246
step: 60, loss: 0.1196683719754219
step: 70, loss: 0.008204689249396324
step: 80, loss: 0.02291327528655529
step: 90, loss: 0.00970697682350874
step: 100, loss: 0.016574684530496597
step: 110, loss: 0.06966487318277359
step: 120, loss: 0.013650134205818176
step: 130, loss: 0.164642333984375
step: 140, loss: 0.045285362750291824
step: 150, loss: 0.008860521018505096
step: 160, loss: 0.10857465118169785
step: 170, loss: 0.09959159046411514
step: 180, loss: 0.004859778564423323
step: 190, loss: 0.07708175480365753
step: 200, loss: 0.007122865878045559
step: 210, loss: 0.037601739168167114
step: 220, loss: 0.0007195539074018598
step: 230, loss: 0.17554421722888947
/usr/local/lib/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=0.9717514124293786, f1=0.9692832764505119,
best_f1=0.9692832764505119
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010342473164200783
step: 10, loss: 0.020989621058106422
step: 20, loss: 0.01991238072514534
step: 30, loss: 0.028828037902712822
step: 40, loss: 0.030171172693371773
step: 50, loss: 0.01925213448703289
step: 60, loss: 0.01623339205980301
step: 70, loss: 0.0008191907545551658
step: 80, loss: 0.0010486090322956443
step: 90, loss: 0.00048213719855993986
step: 100, loss: 0.0028867674991488457
step: 110, loss: 0.015959467738866806
step: 120, loss: 0.01997246779501438
step: 130, loss: 0.0007165411952883005
step: 140, loss: 0.0014893763000145555
step: 150, loss: 0.007271651178598404
step: 160, loss: 0.012041283771395683
step: 170, loss: 0.0011163661256432533
step: 180, loss: 0.002453566761687398
step: 190, loss: 0.022018255665898323
step: 200, loss: 0.002175247063860297
step: 210, loss: 0.004550118464976549
step: 220, loss: 0.0022017196752130985
step: 230, loss: 0.03384935110807419
/usr/local/lib/python3.7/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=0.9651416122004357, f1=0.9616648411829135,
best_f1=0.9692832764505119
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008255647262558341
step: 10, loss: 0.0005913946661166847
step: 20, loss: 0.0019255843944847584
step: 30, loss: 0.00028038170421496034
step: 40, loss: 0.02923687919974327
```

```
step: 50, loss: 0.00414813170209527
step: 60, loss: 0.0007002725033089519
step: 70, loss: 0.0034014927223324776
step: 80, loss: 0.022541899234056473
step: 90, loss: 0.1551647186279297
step: 100, loss: 0.004664591047912836
step: 110, loss: 0.005687100812792778
step: 120, loss: 0.025283081457018852
step: 130, loss: 0.0025025152135640383
step: 140, loss: 0.0047456566244363785
step: 150, loss: 0.0004642292915377766
step: 160, loss: 0.08452475816011429
step: 170, loss: 0.006548622157424688
step: 180, loss: 0.10969379544258118
step: 190, loss: 0.0067677972838282585
step: 200, loss: 0.0029307869262993336
step: 210, loss: 0.0015101036988198757
step: 220, loss: 0.03269356116652489
step: 230, loss: 0.0004594948550220579
/usr/local/lib/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=0.9722530521642618, f1=0.968609865470852,
best_f1=0.968609865470852
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.045681692659855489
step: 10, loss: 0.0071528768166889968
step: 20, loss: 0.0008876023930497468
step: 30, loss: 0.011012932285666466
step: 40, loss: 0.0011533291544765234
step: 50, loss: 0.0002543512382544577
step: 60, loss: 0.0003088160010520369
step: 70, loss: 0.00022674084175378084
step: 80, loss: 0.001827168627642095
step: 90, loss: 0.0004980132798664272
step: 100, loss: 0.010091563686728477
step: 110, loss: 0.0005698202294297516
step: 120, loss: 0.0675126314163208
step: 130, loss: 0.007364946883171797
step: 140, loss: 0.00019204147974960506
step: 150, loss: 0.005459423176944256
step: 160, loss: 0.07894052565097809
```

```
step: 170, loss: 0.09478235244750977
step: 180, loss: 0.08195536583662033
step: 190, loss: 0.00030636307201348245
step: 200, loss: 0.0041085947304964066
step: 210, loss: 0.0004315483383834362
step: 220, loss: 0.009645291604101658
step: 230, loss: 0.0010685963789001107
/usr/local/lib/python3.7/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=0.9704545454545453, f1=0.9657534246575342,
best_f1=0.968609865470852
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007320957258343697
step: 10, loss: 0.000352820148691535
step: 20, loss: 0.0004010794800706208
step: 30, loss: 0.00031292744097299874
step: 40, loss: 0.0004650783084798604
step: 50, loss: 0.029741140082478523
step: 60, loss: 0.0011493571801111102
step: 70, loss: 0.0004169034364167601
step: 80, loss: 0.001190819893963635
step: 90, loss: 0.00019684316066559404
step: 100, loss: 0.0006179865449666977
step: 110, loss: 0.04946840927004814
step: 120, loss: 0.0023812621366232634
step: 130, loss: 0.00035263493191450834
step: 140, loss: 0.0008123043808154762
step: 150, loss: 0.00037584779784083366
step: 160, loss: 0.0014724740758538246
step: 170, loss: 0.0009009978966787457
step: 180, loss: 0.00018821733829099685
step: 190, loss: 0.022242726758122444
step: 200, loss: 0.00019269122276455164
step: 210, loss: 0.0004637185775209218
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 8192.0
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 4096.0
step: 220, loss: 0.00036187784280627966
step: 230, loss: 0.0001531003217678517
/usr/local/lib/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=0.9774266365688488, f1=0.9692832764505119,
best_f1=0.9692832764505119
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.043573617935180664
step: 10, loss: 0.0001245813473360613
step: 20, loss: 0.00028948503313586116
step: 30, loss: 0.0001880656200228259
step: 40, loss: 0.008803142234683037
step: 50, loss: 0.00021523235773202032
step: 60, loss: 0.018398882821202278
step: 70, loss: 0.0011313437717035413
step: 80, loss: 0.0017683164915069938
step: 90, loss: 0.00031129163107834756
step: 100, loss: 0.0008642666507512331
step: 110, loss: 0.0007700322312302887
step: 120, loss: 0.0003938078589271754
step: 130, loss: 0.026805022731423378
step: 140, loss: 0.001015080721117556
step: 150, loss: 0.00013638877135235816
step: 160, loss: 0.00010207537707174197
step: 170, loss: 0.00010897361062234268
step: 180, loss: 0.0003448772768024355
step: 190, loss: 0.0002515691739972681
step: 200, loss: 0.0030586724169552326
step: 210, loss: 0.05064098536968231
step: 220, loss: 0.0002957024844363332
step: 230, loss: 0.0010921047069132328
/usr/local/lib/python3.7/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=0.9683257918552037, f1=0.9681818181818181,
best_f1=0.9692832764505119
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.017497960478067398
step: 10, loss: 0.02016899548470974
step: 20, loss: 0.0009388226317241788
step: 30, loss: 0.00023422205413226038
step: 40, loss: 0.00046026622294448316
step: 50, loss: 0.01647346466779709
step: 60, loss: 0.0009311004541814327
step: 70, loss: 0.00047745267511345446
step: 80, loss: 0.0005261673941276968
step: 90, loss: 0.00013017178571317345
step: 100, loss: 0.00028860237216576934
```

```
step: 110, loss: 0.00056789506925269996
step: 120, loss: 0.00026227973285131156
step: 130, loss: 0.06321340054273605
step: 140, loss: 0.003905740100890398
step: 150, loss: 0.0036368174478411674
step: 160, loss: 0.00046321743866428733
step: 170, loss: 0.00013595697237178683
step: 180, loss: 0.0013972397428005934
step: 190, loss: 6.666184344794601e-05
step: 200, loss: 0.0012592484708875418
step: 210, loss: 7.830945105524734e-05
step: 220, loss: 0.0003364429867360741
step: 230, loss: 0.01165031362324953
/usr/local/lib/python3.7/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=0.9717514124293786, f1=0.9642058165548099,
best_f1=0.9692832764505119
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.61928056110628e-05
step: 10, loss: 0.00019584069377742708
step: 20, loss: 0.003074036445468664
step: 30, loss: 0.00019068876281380653
step: 40, loss: 6.701546953991055e-05
step: 50, loss: 0.0017663240432739258
step: 60, loss: 0.00011286233348073438
step: 70, loss: 0.00016469792171847075
step: 80, loss: 0.07771419733762741
step: 90, loss: 0.024936283007264137
step: 100, loss: 0.0003369305923115462
step: 110, loss: 0.000569167488720268
step: 120, loss: 0.0014942717971280217
step: 130, loss: 0.0006292868056334555
step: 140, loss: 0.0033468659967184067
step: 150, loss: 0.00010209911124547943
step: 160, loss: 0.000371790025383234
step: 170, loss: 6.259901419980451e-05
step: 180, loss: 0.00029134302167221904
step: 190, loss: 7.274506788235158e-05
step: 200, loss: 0.00010664394358173013
step: 210, loss: 0.00015103307669050992
step: 220, loss: 0.025855015963315964
step: 230, loss: 0.026528995484113693
/usr/local/lib/python3.7/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=0.9764309764309763, f1=0.967670011148272,
best_f1=0.9692832764505119
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008945954032242298
step: 10, loss: 0.00013343436876311898
step: 20, loss: 0.00023243714531417936
step: 30, loss: 0.001943072653375566
step: 40, loss: 0.0001978681975742802
step: 50, loss: 0.0008106480818241835
step: 60, loss: 0.00024751623277552426
step: 70, loss: 0.0005343835800886154
step: 80, loss: 0.00024610402761027217
step: 90, loss: 0.00018754773191176355
step: 100, loss: 0.00020167061302345246
step: 110, loss: 0.0005441093817353249
step: 120, loss: 0.01400141790509224
step: 130, loss: 0.00024450934142805636
step: 140, loss: 0.01025464292615652
step: 150, loss: 0.00019234670617152005
step: 160, loss: 0.0009165886440314353
step: 170, loss: 0.0030805584974586964
step: 180, loss: 0.0006950658280402422
step: 190, loss: 0.0003657250781543553
step: 200, loss: 0.00019664267892949283
step: 210, loss: 0.00012822322605643421
step: 220, loss: 0.0006324393907561898
step: 230, loss: 0.0001225801242981106
/usr/local/lib/python3.7/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=0.967816091954023, f1=0.960919540229885,
best_f1=0.9692832764505119
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0010168163571506739
step: 10, loss: 0.0001764857879607007
step: 20, loss: 0.00019733862427528948
step: 30, loss: 0.003253050148487091
step: 40, loss: 0.0010535051114857197
step: 50, loss: 0.0002587676281109452
step: 60, loss: 0.00019639445235952735
step: 70, loss: 0.0003642791125457734
step: 80, loss: 0.022416753694415092
step: 90, loss: 0.00011439065565355122
step: 100, loss: 0.0002961480349767953
```

```
step: 110, loss: 7.803313928889111e-05
step: 120, loss: 0.0003329305036459118
step: 130, loss: 8.697158773429692e-05
step: 140, loss: 9.134867286775261e-05
step: 150, loss: 7.434211147483438e-05
step: 160, loss: 0.00014240120071917772
step: 170, loss: 0.015913913026452065
step: 180, loss: 0.002133206697180867
step: 190, loss: 0.0006181501084938645
step: 200, loss: 9.995193977374583e-05
step: 210, loss: 5.829624205944128e-05
step: 220, loss: 0.00010778279829537496
step: 230, loss: 0.020570049062371254
/usr/local/lib/python3.7/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=0.9725400457665903, f1=0.9692132269099202,
best_f1=0.9692832764505119
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005360727082006633
step: 10, loss: 0.009734324179589748
step: 20, loss: 0.00012383754074107856
step: 30, loss: 0.0003043217584490776
step: 40, loss: 8.5668740211986e-05
step: 50, loss: 0.00011376035399734974
step: 60, loss: 0.0014036301290616393
step: 70, loss: 8.478956442559138e-05
step: 80, loss: 0.00011083375284215435
step: 90, loss: 7.265230669872835e-05
step: 100, loss: 0.00011798907507909462
step: 110, loss: 0.02164800837635994
step: 120, loss: 0.00010515658505028114
step: 130, loss: 0.00047696538968011737
step: 140, loss: 0.0002155550755560398
step: 150, loss: 0.00039677569293417037
step: 160, loss: 0.0027028352487832308
step: 170, loss: 0.00010079027561005205
step: 180, loss: 8.554793748771772e-05
step: 190, loss: 6.567966192960739e-05
step: 200, loss: 0.0008470629109069705
step: 210, loss: 0.00010585062409518287
step: 220, loss: 0.0003305162244942039
step: 230, loss: 0.003859009593725204
/usr/local/lib/python3.7/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=0.9730337078651685, f1=0.9707207207207207,
best_f1=0.9692832764505119
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.956710760481656e-05
step: 10, loss: 8.45703252707608e-05
step: 20, loss: 0.00010371749522164464
step: 30, loss: 0.0002568837662693113
step: 40, loss: 0.00036152213579043746
step: 50, loss: 0.0002757285546977073
step: 60, loss: 0.00013439462054520845
step: 70, loss: 7.841759361326694e-05
step: 80, loss: 0.00020876598136965185
step: 90, loss: 0.0002788228157442063
step: 100, loss: 0.0011401857482269406
step: 110, loss: 0.00010261332499794662
step: 120, loss: 6.881565059302375e-05
step: 130, loss: 0.00013166472490411252
step: 140, loss: 0.00011378740600775927
step: 150, loss: 0.00014058142551220953
step: 160, loss: 6.282459798967466e-05
step: 170, loss: 6.331045733531937e-05
step: 180, loss: 0.0003601904318202287
step: 190, loss: 5.539977064472623e-05
step: 200, loss: 8.334004814969376e-05
step: 210, loss: 0.00016017432790249586
step: 220, loss: 4.881689164903946e-05
step: 230, loss: 4.664199150283821e-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=0.9750566893424036, f1=0.9705882352941176,
best_f1=0.9692832764505119
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.650365634821355e-05
step: 10, loss: 7.874649600125849e-05
step: 20, loss: 0.00025511422427371144
step: 30, loss: 0.0003823733713943511
step: 40, loss: 7.432346319546923e-05
step: 50, loss: 7.892218854976818e-05
step: 60, loss: 4.672251452575438e-05
step: 70, loss: 7.993623148649931e-05
step: 80, loss: 8.662602340336889e-05
step: 90, loss: 9.806223533814773e-05
step: 100, loss: 0.0031543441582471132
```

```
step: 110, loss: 0.0006376567762345076
step: 120, loss: 0.008804451674222946
step: 130, loss: 9.312728070653975e-05
step: 140, loss: 5.9377918660175055e-05
step: 150, loss: 0.00014296693552751094
step: 160, loss: 7.394857675535604e-05
step: 170, loss: 8.056403748923913e-05
step: 180, loss: 0.000137229886604473
step: 190, loss: 0.0004865280934609473
step: 200, loss: 4.574049307848327e-05
step: 210, loss: 0.0020432595629245043
step: 220, loss: 9.504299669060856e-05
step: 230, loss: 2.8870339519926347e-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=0.976324689966178, f1=0.971815107102593,
best_f1=0.9692832764505119
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00026990043252505362
step: 10, loss: 9.047111234394833e-05
step: 20, loss: 0.00013319196295924485
step: 30, loss: 0.00013353924441616982
step: 40, loss: 0.002006364520639181
step: 50, loss: 0.00014816429757047445
step: 60, loss: 5.514195072464645e-05
step: 70, loss: 0.0008478689123876393
step: 80, loss: 6.724436389049515e-05
step: 90, loss: 3.5120563552482054e-05
step: 100, loss: 9.193797450279817e-05
step: 110, loss: 5.842695463798009e-05
step: 120, loss: 0.00018077658023685217
step: 130, loss: 7.100868242559955e-05
step: 140, loss: 0.00010097958875121549
step: 150, loss: 9.13442563614808e-05
step: 160, loss: 4.154968337388709e-05
step: 170, loss: 4.0186198020819575e-05
step: 180, loss: 5.5517088185297325e-05
step: 190, loss: 0.011469007469713688
step: 200, loss: 9.634863818064332e-05
step: 210, loss: 0.02523796446621418
step: 220, loss: 0.0001206581509904936
step: 230, loss: 8.592524682171643e-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=0.9752252252252253, f1=0.971815107102593,
best_f1=0.9692832764505119
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, 237.20it/s]
load_f1 = 0.9739524348810873
real_f1 = 0.9774266365688488
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 [00:16, 261.20it/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 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: 0.7957703471183777
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.45284727215766907
step: 20, loss: 0.49903640151023865
step: 30, loss: 0.464160680770874
step: 40, loss: 0.4749623239040375
step: 50, loss: 0.4543660879135132
step: 60, loss: 0.5257434248924255
step: 70, loss: 0.21057423949241638
step: 80, loss: 0.18718047440052032
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 90, loss: 0.2569151222705841
step: 100, loss: 0.15560761094093323
step: 110, loss: 0.11373285949230194
step: 120, loss: 0.10693087428808212
step: 130, loss: 0.052636709064245224
step: 140, loss: 0.2502637803554535
step: 150, loss: 0.07146061956882477
step: 160, loss: 0.17828932404518127
step: 170, loss: 0.3813953697681427
step: 180, loss: 0.18677741289138794
step: 190, loss: 0.0780547708272934
step: 200, loss: 0.09362553060054779
step: 210, loss: 0.10203146934509277
step: 220, loss: 0.12706942856311798
step: 230, loss: 0.06848398596048355
step: 240, loss: 0.11413244903087616
step: 250, loss: 0.10981154441833496
step: 260, loss: 0.027040736749768257
step: 270, loss: 0.05673886835575104
step: 280, loss: 0.12142543494701385
step: 290, loss: 0.16390518844127655
step: 300, loss: 0.1296723186969757
step: 310, loss: 0.13384319841861725
```

```
step: 320, loss: 0.20265375077724457
step: 330, loss: 0.19623322784900665
step: 340, loss: 0.18340131640434265
step: 350, loss: 0.07456284761428833
step: 360, loss: 0.0732649713754654
step: 370, loss: 0.15772028267383575
step: 380, loss: 0.1728438287973404
step: 390, loss: 0.23406672477722168
step: 400, loss: 0.03752248361706734
step: 410, loss: 0.09542381763458252
step: 420, loss: 0.053496211767196655
step: 430, loss: 0.11102361977100372
step: 440, loss: 0.22999562323093414
step: 450, loss: 0.05922868102788925
step: 460, loss: 0.05790559947490692
step: 470, loss: 0.21066737174987793
step: 480, loss: 0.21064049005508423
step: 490, loss: 0.06794645637273788
step: 500, loss: 0.022641710937023163
step: 510, loss: 0.05514945462346077
step: 520, loss: 0.06730180233716965
step: 530, loss: 0.21041423082351685
/usr/local/lib/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=0.8935969868173258, f1=0.8918156161806209,
best_f1=0.8918156161806209
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1963111162185669
step: 10, loss: 0.1030094102025032
step: 20, loss: 0.09878762066364288
step: 30, loss: 0.08026096969842911
step: 40, loss: 0.032393842935562134
step: 50, loss: 0.02487899363040924
step: 60, loss: 0.0930054634809494
step: 70, loss: 0.23025423288345337
step: 80, loss: 0.030764468014240265
step: 90, loss: 0.03229852393269539
step: 100, loss: 0.2747279405593872
step: 110, loss: 0.07866419851779938
step: 120, loss: 0.06940712779760361
step: 130, loss: 0.06262057274580002
```

```
step: 140, loss: 0.019640129059553146
step: 150, loss: 0.05670788511633873
step: 160, loss: 0.10363828390836716
step: 170, loss: 0.07501628249883652
step: 180, loss: 0.07787946611642838
step: 190, loss: 0.055625706911087036
step: 200, loss: 0.08521855622529984
step: 210, loss: 0.0995684266090393
step: 220, loss: 0.099794901907444
step: 230, loss: 0.1299094706773758
step: 240, loss: 0.10784344375133514
step: 250, loss: 0.08131592720746994
step: 260, loss: 0.02262706682085991
step: 270, loss: 0.07827462255954742
step: 280, loss: 0.08065181970596313
step: 290, loss: 0.0597449466586113
step: 300, loss: 0.025014888495206833
step: 310, loss: 0.044829919934272766
step: 320, loss: 0.16546858847141266
step: 330, loss: 0.04514255002140999
step: 340, loss: 0.08158759772777557
step: 350, loss: 0.08476370573043823
step: 360, loss: 0.09433554857969284
step: 370, loss: 0.03601652756333351
step: 380, loss: 0.09262759238481522
step: 390, loss: 0.08177245408296585
step: 400, loss: 0.08067094534635544
step: 410, loss: 0.01656956411898136
step: 420, loss: 0.0654192790389061
step: 430, loss: 0.05056837573647499
step: 440, loss: 0.016609016805887222
step: 450, loss: 0.19314740598201752
step: 460, loss: 0.2406437247991562
step: 470, loss: 0.03825636953115463
step: 480, loss: 0.22861172258853912
step: 490, loss: 0.08342192322015762
step: 500, loss: 0.02450701966881752
step: 510, loss: 0.1610947847366333
step: 520, loss: 0.040629882365465164
step: 530, loss: 0.18069520592689514
/usr/local/lib/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=0.9060402684563759, f1=0.8983700862895493,
```

```
best_f1=0.8983700862895493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.026418542489409447
step: 10, loss: 0.16480779647827148
step: 20, loss: 0.04013039916753769
step: 30, loss: 0.2535559833049774
step: 40, loss: 0.0194233488291502
step: 50, loss: 0.018177587538957596
step: 60, loss: 0.024339376017451286
step: 70, loss: 0.012224413454532623
step: 80, loss: 0.06874936819076538
step: 90, loss: 0.2440304458141327
step: 100, loss: 0.009478501975536346
step: 110, loss: 0.011974828317761421
step: 120, loss: 0.008110460825264454
step: 130, loss: 0.00677523436024785
step: 140, loss: 0.15567344427108765
step: 150, loss: 0.050797611474990845
step: 160, loss: 0.06841482222080231
step: 170, loss: 0.12403298169374466
step: 180, loss: 0.10727682709693909
step: 190, loss: 0.009281753562390804
step: 200, loss: 0.09812574833631516
step: 210, loss: 0.1079275980591774
step: 220, loss: 0.018675226718187332
step: 230, loss: 0.0064268559217453
step: 240, loss: 0.009542834013700485
step: 250, loss: 0.029279964044690132
step: 260, loss: 0.015136636793613434
step: 270, loss: 0.012293776497244835
step: 280, loss: 0.13606972992420197
step: 290, loss: 0.0661509782075882
step: 300, loss: 0.0434640571475029
step: 310, loss: 0.0975421667098999
step: 320, loss: 0.013728848658502102
step: 330, loss: 0.015777939930558205
step: 340, loss: 0.010419540107250214
step: 350, loss: 0.0661400556564331
step: 360, loss: 0.008326717652380466
step: 370, loss: 0.018391352146863937
step: 380, loss: 0.02224419079720974
step: 390, loss: 0.05845740810036659
step: 400, loss: 0.06091609597206116
step: 410, loss: 0.1221604123711586
step: 420, loss: 0.1062026172876358
step: 430, loss: 0.047148317098617554
```

```
step: 440, loss: 0.13498428463935852
step: 450, loss: 0.2002817541360855
step: 460, loss: 0.10943828523159027
step: 470, loss: 0.01735713705420494
step: 480, loss: 0.04762881249189377
step: 490, loss: 0.04348323121666908
step: 500, loss: 0.023896627128124237
step: 510, loss: 0.008806592784821987
step: 520, loss: 0.07922468334436417
step: 530, loss: 0.00988561287522316
/usr/local/lib/python3.7/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=0.9032558139534883, f1=0.9028625058657908,
best_f1=0.8983700862895493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007284555584192276
step: 10, loss: 0.0026664421893656254
step: 20, loss: 0.022051256150007248
step: 30, loss: 0.004081401042640209
step: 40, loss: 0.019137412309646606
step: 50, loss: 0.03605726733803749
step: 60, loss: 0.005476204212754965
step: 70, loss: 0.004785778000950813
step: 80, loss: 0.006453245412558317
step: 90, loss: 0.019534483551979065
step: 100, loss: 0.08591578900814056
step: 110, loss: 0.0022644998971372843
step: 120, loss: 0.00463572284206748
step: 130, loss: 0.04559957981109619
step: 140, loss: 0.03538719564676285
step: 150, loss: 0.008724336512386799
step: 160, loss: 0.0045746006071567535
step: 170, loss: 0.002844797680154443
step: 180, loss: 0.011919708922505379
step: 190, loss: 0.0010054142912849784
step: 200, loss: 0.01815889962017536
step: 210, loss: 0.15489964187145233
step: 220, loss: 0.06370159238576889
step: 230, loss: 0.44903379678726196
step: 240, loss: 0.015547635033726692
step: 250, loss: 0.004893516656011343
step: 260, loss: 0.1590399444103241
step: 270, loss: 0.14315423369407654
step: 280, loss: 0.0014167819172143936
step: 290, loss: 0.1046871691942215
```

```
step: 300, loss: 0.015232881531119347
step: 310, loss: 0.004540650174021721
step: 320, loss: 0.008483648300170898
step: 330, loss: 0.08558277040719986
step: 340, loss: 0.08373722434043884
step: 350, loss: 0.021801909431815147
step: 360, loss: 0.12851890921592712
step: 370, loss: 0.024300893768668175
step: 380, loss: 0.008220350369811058
step: 390, loss: 0.009879604913294315
step: 400, loss: 0.018951255828142166
step: 410, loss: 0.02749033086001873
step: 420, loss: 0.033634815365076065
step: 430, loss: 0.005339920520782471
step: 440, loss: 0.04988139495253563
step: 450, loss: 0.03003094717860222
step: 460, loss: 0.0012233840534463525
step: 470, loss: 0.049025651067495346
step: 480, loss: 0.0069290874525904655
step: 490, loss: 0.03689606860280037
step: 500, loss: 0.012131918221712112
step: 510, loss: 0.15430089831352234
step: 520, loss: 0.1280411332845688
step: 530, loss: 0.002400639932602644
/usr/local/lib/python3.7/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=0.9011600928074246, f1=0.8970656730321379,
best_f1=0.8983700862895493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.011557993479073048
step: 10, loss: 0.012165124528110027
step: 20, loss: 0.004259068053215742
step: 30, loss: 0.003912209067493677
step: 40, loss: 0.10309725999832153
step: 50, loss: 0.020277656614780426
step: 60, loss: 0.005290071479976177
step: 70, loss: 0.0008052595076151192
step: 80, loss: 0.0008398643112741411
step: 90, loss: 0.01754801906645298
step: 100, loss: 0.0008760356577113271
step: 110, loss: 0.004788028076291084
step: 120, loss: 0.05739276483654976
step: 130, loss: 0.0013131623854860663
step: 140, loss: 0.0019485818920657039
step: 150, loss: 0.0017627287888899446
```

```
step: 160, loss: 0.04007219523191452
step: 170, loss: 0.006592475343495607
step: 180, loss: 0.0011561164865270257
step: 190, loss: 0.005392597988247871
step: 200, loss: 0.034529201686382294
step: 210, loss: 0.00539179053157568
step: 220, loss: 0.0007385583594441414
step: 230, loss: 0.000662448292132467
step: 240, loss: 0.00981524121016264
step: 250, loss: 0.00039553618989884853
step: 260, loss: 0.006691875867545605
step: 270, loss: 0.02896358259022236
step: 280, loss: 0.14538820087909698
step: 290, loss: 0.002648907946422696
step: 300, loss: 0.01401901338249445
step: 310, loss: 0.0043207574635744095
step: 320, loss: 0.08957386016845703
step: 330, loss: 0.00210730847902596
step: 340, loss: 0.0038970480673015118
step: 350, loss: 0.005027954000979662
step: 360, loss: 0.15959274768829346
step: 370, loss: 0.0033750219736248255
step: 380, loss: 0.02418803982436657
step: 390, loss: 0.04768470674753189
step: 400, loss: 0.09555517882108688
step: 410, loss: 0.0005290394183248281
step: 420, loss: 0.0015727889258414507
step: 430, loss: 0.0027230659034103155
step: 440, loss: 0.10737797617912292
step: 450, loss: 0.0036282793153077364
step: 460, loss: 0.0013130366569384933
step: 470, loss: 0.09026187658309937
step: 480, loss: 0.011017585173249245
step: 490, loss: 0.012628080323338509
step: 500, loss: 0.03693442419171333
step: 510, loss: 0.031206514686346054
step: 520, loss: 0.003847497282549739
step: 530, loss: 0.009805646724998951
/usr/local/lib/python3.7/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=0.9026303645592986, f1=0.8993973110802038,
best_f1=0.8983700862895493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.016490183770656586
step: 10, loss: 0.006425876170396805
```

```
step: 20, loss: 0.000981492456048727
step: 30, loss: 0.000986986793577671
step: 40, loss: 0.0033893270883709192
step: 50, loss: 0.002752600936219096
step: 60, loss: 0.0020274363923817873
step: 70, loss: 0.015113738365471363
step: 80, loss: 0.000458858092315495
step: 90, loss: 0.00027631883858703077
step: 100, loss: 0.07961740344762802
step: 110, loss: 0.002555467654019594
step: 120, loss: 0.011166478507220745
step: 130, loss: 0.0007631295011378825
step: 140, loss: 0.002200909424573183
step: 150, loss: 0.005093386396765709
step: 160, loss: 0.0004060078936163336
step: 170, loss: 0.00023183797020465136
step: 180, loss: 0.000453210726846095
step: 190, loss: 0.03447427973151207
step: 200, loss: 0.00019249532488174736
step: 210, loss: 0.00026034639449790012
step: 220, loss: 0.0011904468992725015
step: 230, loss: 0.05130118131637573
step: 240, loss: 0.034498460590839386
step: 250, loss: 0.007509704679250717
step: 260, loss: 0.016432203352451324
step: 270, loss: 0.03510946035385132
step: 280, loss: 0.023546144366264343
step: 290, loss: 0.07066531479358673
step: 300, loss: 0.024523958563804626
step: 310, loss: 0.0048355092294514148
step: 320, loss: 0.1976853758096695
step: 330, loss: 0.06634149700403214
step: 340, loss: 0.013279481790959835
step: 350, loss: 0.16070085763931274
step: 360, loss: 0.00036723876837641
step: 370, loss: 0.17868678271770477
step: 380, loss: 0.008367855101823807
step: 390, loss: 0.1121731624007225
step: 400, loss: 0.0010164215927943587
step: 410, loss: 0.0046013458631932735
step: 420, loss: 0.009226629510521889
step: 430, loss: 0.0011221240274608135
step: 440, loss: 0.0049110655672848225
step: 450, loss: 0.0009276081109419465
step: 460, loss: 0.20466193556785583
step: 470, loss: 0.004761855583637953
step: 480, loss: 0.013506026938557625
step: 490, loss: 0.004001706838607788
```

```
step: 500, loss: 0.008353087119758129
step: 510, loss: 0.02723833918571472
step: 520, loss: 0.0010374002158641815
step: 530, loss: 0.01313795242458582
/usr/local/lib/python3.7/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=0.8908418131359852, f1=0.8912442396313364,
best_f1=0.8983700862895493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07333767414093018
step: 10, loss: 0.01973562128841877
step: 20, loss: 0.003330631647258997
step: 30, loss: 0.005626286379992962
step: 40, loss: 0.0024563393089920282
step: 50, loss: 0.009741359390318394
step: 60, loss: 0.00026426964905111298
step: 70, loss: 0.00047280528815463185
step: 80, loss: 0.000455594650702551
step: 90, loss: 0.0016369371442124248
step: 100, loss: 0.0021601072512567043
step: 110, loss: 0.0008366628317162395
step: 120, loss: 0.002632472664117813
step: 130, loss: 0.00021598349849227816
step: 140, loss: 0.0009831400820985436
step: 150, loss: 0.00019168941071256995
step: 160, loss: 0.0005799129721708596
step: 170, loss: 0.0004290013457648456
step: 180, loss: 0.001221152488142252
step: 190, loss: 0.0004218678514007479
step: 200, loss: 0.05612535774707794
step: 210, loss: 0.0014962140703573823
step: 220, loss: 0.0015119112795218825
step: 230, loss: 0.0008806049008853734
step: 240, loss: 0.006690976209938526
step: 250, loss: 0.0008193666581064463
step: 260, loss: 0.001178559847176075
step: 270, loss: 0.00017111508350353688
step: 280, loss: 0.010083788074553013
step: 290, loss: 0.005692333914339542
step: 300, loss: 0.0005297386669553816
step: 310, loss: 0.0004740360891446471
step: 320, loss: 0.0039343442767858505
step: 330, loss: 0.0008625040063634515
step: 340, loss: 0.18687357008457184
step: 350, loss: 0.07974398881196976
```

```
step: 360, loss: 0.006429252214729786
step: 370, loss: 0.0004858885076828301
step: 380, loss: 0.0019650179892778397
step: 390, loss: 0.0007177767110988498
step: 400, loss: 0.0634099468588829
step: 410, loss: 0.0006031243246980011
step: 420, loss: 0.0005333191948011518
step: 430, loss: 0.00021373962226789445
step: 440, loss: 0.00045395290362648666
step: 450, loss: 0.0018471889197826385
step: 460, loss: 0.0027088383212685585
step: 470, loss: 0.00262046093121171
step: 480, loss: 0.0019147660350427032
step: 490, loss: 0.003951144404709339
step: 500, loss: 0.00052813929505564623
step: 510, loss: 0.00796863529831171
step: 520, loss: 0.069069042801857
step: 530, loss: 0.02165716327726841
/usr/local/lib/python3.7/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=0.8995835261453031, f1=0.9022140221402214,
best_f1=0.8983700862895493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012072538025677204
step: 10, loss: 0.014736533164978027
step: 20, loss: 0.0010048802942037582
step: 30, loss: 0.0004144013801123947
step: 40, loss: 0.003550904802978039
step: 50, loss: 0.0027104048058390617
step: 60, loss: 0.008827907033264637
step: 70, loss: 0.07480009645223618
step: 80, loss: 0.000330368522554636
step: 90, loss: 0.0006908816867507994
step: 100, loss: 0.0009387299069203436
step: 110, loss: 0.0017630885122343898
step: 120, loss: 0.002282259054481983
step: 130, loss: 0.000237695625401102
step: 140, loss: 0.004948015790432692
step: 150, loss: 0.00044532708125188947
step: 160, loss: 0.0003884357283823192
step: 170, loss: 0.00017414231842849404
step: 180, loss: 0.00016018351016100496
step: 190, loss: 0.02655237913131714
step: 200, loss: 0.0003371702623553574
step: 210, loss: 0.17466062307357788
```

```
step: 220, loss: 0.006897317245602608
step: 230, loss: 0.00022241054102778435
step: 240, loss: 0.012574972584843636
step: 250, loss: 0.0003925265045836568
step: 260, loss: 0.03526874631643295
step: 270, loss: 0.0002336531615583226
step: 280, loss: 0.00029168519540689886
step: 290, loss: 0.0005458337836898863
step: 300, loss: 0.00020137522369623184
step: 310, loss: 0.012629357166588306
step: 320, loss: 0.0015925532206892967
step: 330, loss: 0.0007533750031143427
step: 340, loss: 0.00011560953862499446
step: 350, loss: 0.0016007270896807313
step: 360, loss: 0.01443071011453867
step: 370, loss: 0.004994941409677267
step: 380, loss: 0.006644176784902811
step: 390, loss: 0.0009287333232350647
step: 400, loss: 0.11687012016773224
step: 410, loss: 0.053381383419036865
step: 420, loss: 0.0019245248986408114
step: 430, loss: 0.0006322598783299327
step: 440, loss: 0.00026368367252871394
step: 450, loss: 0.013609025627374649
step: 460, loss: 0.04697421193122864
step: 470, loss: 0.0002699576143641025
step: 480, loss: 0.0003249058499932289
step: 490, loss: 0.0007191228214651346
step: 500, loss: 0.0020133040379732847
step: 510, loss: 0.00011976118548773229
step: 520, loss: 0.0001653080980759114
step: 530, loss: 0.0001383156341034919
/usr/local/lib/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=0.9060925332111772, f1=0.8996350364963502,
best_f1=0.8996350364963502
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0036786366254091263
step: 10, loss: 9.784350550035015e-05
step: 20, loss: 0.0015226822579279542
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
```

```
step: 30, loss: 0.000552757759578526
step: 40, loss: 0.000308136863168329
step: 50, loss: 0.00016828106890898198
step: 60, loss: 0.002017783699557185
step: 70, loss: 0.014633704908192158
step: 80, loss: 8.836078632157296e-05
step: 90, loss: 0.0026207466144114733
step: 100, loss: 0.0005560661084018648
step: 110, loss: 0.00820793304592371
step: 120, loss: 0.0002566213661339134
step: 130, loss: 0.0009100298047997057
step: 140, loss: 0.007589375600218773
step: 150, loss: 0.0005159458960406482
step: 160, loss: 0.0001651107450015843
step: 170, loss: 0.0006832600338384509
step: 180, loss: 0.0002207651996286586
step: 190, loss: 0.0009102042531594634
step: 200, loss: 0.0001870456908363849
step: 210, loss: 0.00023078967933543026
step: 220, loss: 6.792187923565507e-05
step: 230, loss: 0.0001397984888171777
step: 240, loss: 0.00038889984716661274
step: 250, loss: 0.00018294030451215804
step: 260, loss: 5.7903969718609005e-05
step: 270, loss: 0.00015331456961575896
step: 280, loss: 6.260956433834508e-05
step: 290, loss: 5.234386117081158e-05
step: 300, loss: 0.00020979686814825982
step: 310, loss: 4.88580008095596e-05
step: 320, loss: 6.823305739089847e-05
step: 330, loss: 0.00010147925786441192
step: 340, loss: 0.00049297243822366
step: 350, loss: 5.4649764933856204e-05
step: 360, loss: 0.00439056009054184
step: 370, loss: 0.0010298638371750712
step: 380, loss: 7.582598482258618e-05
step: 390, loss: 0.00030572060495615005
step: 400, loss: 0.0003216903714928776
step: 410, loss: 0.014908099547028542
step: 420, loss: 0.0014983774162828922
step: 430, loss: 4.7141333197942004e-05
step: 440, loss: 0.00010858186578843743
step: 450, loss: 0.0009963561315089464
step: 460, loss: 7.947564881760627e-05
step: 470, loss: 6.476209091488272e-05
step: 480, loss: 0.00052045495249331
step: 490, loss: 6.791246414650232e-05
step: 500, loss: 5.4684263886883855e-05
```

```
step: 510, loss: 0.0004167479055467993
step: 520, loss: 0.012371503747999668
step: 530, loss: 5.751325443270616e-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=0.9033749422098938, f1=0.901669758812616,
best_f1=0.8996350364963502
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0050480980426073074
step: 10, loss: 4.822778646484949e-05
step: 20, loss: 6.349948671413586e-05
step: 30, loss: 5.258915189187974e-05
step: 40, loss: 0.00010075852333102375
step: 50, loss: 6.001077781547792e-05
step: 60, loss: 6.108714296715334e-05
step: 70, loss: 0.00015450469800271094
step: 80, loss: 0.0007995122577995062
step: 90, loss: 8.860880916472524e-05
step: 100, loss: 0.0020601702854037285
step: 110, loss: 0.00028528974507935345
step: 120, loss: 0.006761244963854551
step: 130, loss: 0.00015482779417652637
step: 140, loss: 0.0008931452175602317
step: 150, loss: 6.86238709022291e-05
step: 160, loss: 0.00016742282605264336
step: 170, loss: 0.00433776481077075
step: 180, loss: 0.0009750283788889647
step: 190, loss: 0.0003633436281234026
step: 200, loss: 0.0017439196817576885
step: 210, loss: 7.74811051087454e-05
step: 220, loss: 0.0016206064028665423
step: 230, loss: 0.000104499478766229
step: 240, loss: 0.0003934820124413818
step: 250, loss: 0.002161746146157384
step: 260, loss: 0.00028509151889011264
step: 270, loss: 8.515421359334141e-05
step: 280, loss: 0.0002393585309619084
step: 290, loss: 6.072978794691153e-05
step: 300, loss: 0.001805689069442451
step: 310, loss: 0.0002200217277277261
step: 320, loss: 4.801155591849238e-05
step: 330, loss: 0.0001225253363372758
step: 340, loss: 0.00010765276965685189
step: 350, loss: 6.873391976114362e-05
step: 360, loss: 8.59339561429806e-05
```

```
step: 370, loss: 0.0011056307703256607
step: 380, loss: 6.449471402447671e-05
step: 390, loss: 0.000187340920092538
step: 400, loss: 0.0003122298512607813
step: 410, loss: 8.079632971202955e-05
step: 420, loss: 0.00019213106133975089
step: 430, loss: 0.00015844847075641155
step: 440, loss: 7.898172043496743e-05
step: 450, loss: 0.019542722031474113
step: 460, loss: 0.00021635439770761877
step: 470, loss: 0.0020542836282402277
step: 480, loss: 7.290548091987148e-05
step: 490, loss: 0.038443200290203094
step: 500, loss: 0.0005690681282430887
step: 510, loss: 0.0010591563768684864
step: 520, loss: 0.006013853941112757
step: 530, loss: 0.000254300597589463
/usr/local/lib/python3.7/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=0.9016470588235295, f1=0.8930581613508444,
best_f1=0.8996350364963502
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007233509095385671
step: 10, loss: 0.000499016314279288
step: 20, loss: 0.0001110262528527528
step: 30, loss: 0.00035420217318460345
step: 40, loss: 0.0028351445216685534
step: 50, loss: 8.716104639461264e-05
step: 60, loss: 0.00014019827358424664
step: 70, loss: 8.283962233690545e-05
step: 80, loss: 6.761603435734287e-05
step: 90, loss: 0.00010156854841625318
step: 100, loss: 0.0006439025746658444
step: 110, loss: 0.0075029293075203896
step: 120, loss: 0.006398208439350128
step: 130, loss: 5.637681533698924e-05
step: 140, loss: 8.7399770563934e-05
step: 150, loss: 0.00019131848239339888
step: 160, loss: 0.00015001185238361359
step: 170, loss: 0.0002498566173017025
step: 180, loss: 0.0010797104332596064
step: 190, loss: 0.0003445987240411341
step: 200, loss: 0.0004644383443519473
step: 210, loss: 0.0005642975447699428
step: 220, loss: 8.744795195525512e-05
```

```
step: 230, loss: 6.75944538670592e-05
step: 240, loss: 0.030654557049274445
step: 250, loss: 5.043044438934885e-05
step: 260, loss: 8.79075814737007e-05
step: 270, loss: 0.00036501933936960995
step: 280, loss: 0.00010021060006693006
step: 290, loss: 0.000524923496413976
step: 300, loss: 0.003069440834224224
step: 310, loss: 0.00942252017557621
step: 320, loss: 0.00031900827889330685
step: 330, loss: 0.0006577761378139257
step: 340, loss: 8.208440704038367e-05
step: 350, loss: 0.00013197668886277825
step: 360, loss: 9.550144022796303e-05
step: 370, loss: 6.796663365093991e-05
step: 380, loss: 5.918038368690759e-05
step: 390, loss: 0.04471724107861519
step: 400, loss: 0.0002497246314305812
step: 410, loss: 7.974840991664678e-05
step: 420, loss: 0.0059918370097875595
step: 430, loss: 0.0043881731107831
step: 440, loss: 0.0005119277047924697
step: 450, loss: 4.4045114918844774e-05
step: 460, loss: 0.007246713619679213
step: 470, loss: 0.0007990517187863588
step: 480, loss: 0.014262843877077103
step: 490, loss: 0.000770953600294888
step: 500, loss: 0.011991828680038452
step: 510, loss: 0.00015751566388644278
step: 520, loss: 0.0005350152496248484
step: 530, loss: 9.69211760093458e-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=0.8988149498632635, f1=0.9036969420356001,
best_f1=0.8996350364963502
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008407214772887528
step: 10, loss: 6.523810588987544e-05
step: 20, loss: 8.323884685523808e-05
step: 30, loss: 0.0002600625157356262
step: 40, loss: 6.419656710932031e-05
step: 50, loss: 3.0509403586620465e-05
step: 60, loss: 0.0007566519198007882
step: 70, loss: 0.00020699625334115699
step: 80, loss: 0.00013380579184740782
```

```
step: 90, loss: 0.00019889949180651456
step: 100, loss: 3.906920028384775e-05
step: 110, loss: 8.144558523781598e-05
step: 120, loss: 0.00013022197526879609
step: 130, loss: 4.339013321441598e-05
step: 140, loss: 4.7305249609053135e-05
step: 150, loss: 0.00015470771177206188
step: 160, loss: 7.83177965786308e-05
step: 170, loss: 2.8944879886694252e-05
step: 180, loss: 5.321920252754353e-05
step: 190, loss: 4.0575203456683084e-05
step: 200, loss: 5.626972779282369e-05
step: 210, loss: 0.00030474993400275707
step: 220, loss: 3.9132577512646094e-05
step: 230, loss: 6.898136052768677e-05
step: 240, loss: 4.5449014578480273e-05
step: 250, loss: 4.484601959120482e-05
step: 260, loss: 4.0998005715664476e-05
step: 270, loss: 3.712440229719505e-05
step: 280, loss: 7.100898073986173e-05
step: 290, loss: 0.00017724283679854125
step: 300, loss: 0.0001684194867266342
step: 310, loss: 4.164713027421385e-05
step: 320, loss: 3.972088234149851e-05
step: 330, loss: 0.0002167819329770282
step: 340, loss: 0.0004768634680658579
step: 350, loss: 0.0015892706578597426
step: 360, loss: 5.209603114053607e-05
step: 370, loss: 3.1566953111905605e-05
step: 380, loss: 5.29929602635093e-05
step: 390, loss: 3.1071969715412706e-05
step: 400, loss: 3.67524080502335e-05
step: 410, loss: 0.0009631288703531027
step: 420, loss: 4.957252895110287e-05
step: 430, loss: 4.1624381992733106e-05
step: 440, loss: 0.00021496166300494224
step: 450, loss: 0.1485215425491333
step: 460, loss: 0.00010233257489744574
step: 470, loss: 0.0001466476242057979
step: 480, loss: 4.297719715395942e-05
step: 490, loss: 0.005113240797072649
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 500, loss: 0.003638913156464696
step: 510, loss: 4.001178967882879e-05
step: 520, loss: 4.484942473936826e-05
step: 530, loss: 0.001627355464734137
/usr/local/lib/python3.7/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=0.8905950095969289, f1=0.8904830224772837,
best_f1=0.8996350364963502
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00010752586968010291
step: 10, loss: 3.423423186177388e-05
step: 20, loss: 0.0007462750072591007
step: 30, loss: 0.0038834363222122192
step: 40, loss: 3.93554764741566e-05
step: 50, loss: 0.00019814104598481208
step: 60, loss: 0.00018984684720635414
step: 70, loss: 3.396561442059465e-05
step: 80, loss: 7.651395571883768e-05
step: 90, loss: 0.0013146733399480581
step: 100, loss: 4.826668373425491e-05
step: 110, loss: 8.044319838518277e-05
step: 120, loss: 0.00020069078891538084
step: 130, loss: 7.151308818720281e-05
step: 140, loss: 0.00018645150703378022
step: 150, loss: 3.154865407850593e-05
step: 160, loss: 0.0008231017272919416
step: 170, loss: 9.206373943015933e-05
step: 180, loss: 0.00027063069865107536
step: 190, loss: 0.0012436588294804096
step: 200, loss: 0.002125989180058241
step: 210, loss: 0.00021776296489406377
step: 220, loss: 0.00011434597399784252
step: 230, loss: 2.851603858289309e-05
step: 240, loss: 9.183723159367219e-05
step: 250, loss: 0.0033737635239958763
step: 260, loss: 0.0002535344974603504
step: 270, loss: 0.00012284166587051004
step: 280, loss: 0.0002059113176073879
step: 290, loss: 0.0005261465557850897
step: 300, loss: 9.351284825243056e-05
step: 310, loss: 2.9465663828887045e-05
step: 320, loss: 0.00017607523477636278
step: 330, loss: 4.7728877689223737e-05
step: 340, loss: 4.219428592477925e-05
step: 350, loss: 5.820185469929129e-05
step: 360, loss: 2.6877252821577713e-05
step: 370, loss: 0.003508162917569279
step: 380, loss: 6.509045488201082e-05
step: 390, loss: 5.130837234901264e-05
step: 400, loss: 2.634815791680012e-05
step: 410, loss: 6.679175567114726e-05
```

```
step: 420, loss: 4.591913602780551e-05
step: 430, loss: 0.00010014799045166001
step: 440, loss: 7.019101758487523e-05
step: 450, loss: 3.907267819158733e-05
step: 460, loss: 2.3003110982244834e-05
step: 470, loss: 0.0011808399576693773
step: 480, loss: 0.00014912246842868626
step: 490, loss: 0.0001069881473085843
step: 500, loss: 8.436105417786166e-05
step: 510, loss: 4.4010812416672707e-05
step: 520, loss: 2.536469946790021e-05
step: 530, loss: 4.984275437891483e-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=0.896358543417367, f1=0.8978644382544103,
best_f1=0.8996350364963502
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.450814882759005e-05
step: 10, loss: 0.00018270683358423412
step: 20, loss: 9.419917478226125e-05
step: 30, loss: 0.00010236304660793394
step: 40, loss: 3.471408854238689e-05
step: 50, loss: 0.00018256819748785347
step: 60, loss: 3.7165485991863534e-05
step: 70, loss: 2.602799577289261e-05
step: 80, loss: 3.621943324105814e-05
step: 90, loss: 2.1412513888208196e-05
step: 100, loss: 4.236587119521573e-05
step: 110, loss: 3.814052979578264e-05
step: 120, loss: 3.9959559217095375e-05
step: 130, loss: 2.510022204660345e-05
step: 140, loss: 0.0005873705958947539
step: 150, loss: 0.0003167249960824847
step: 160, loss: 0.0010453663999214768
step: 170, loss: 0.00022505054948851466
step: 180, loss: 7.15884380042553e-05
step: 190, loss: 0.00020405893155839294
step: 200, loss: 0.00012380695261526853
step: 210, loss: 0.0005059116519987583
step: 220, loss: 0.0001514936884632334
step: 230, loss: 0.00017833923629950732
step: 240, loss: 4.079611608176492e-05
step: 250, loss: 0.00013897671306040138
step: 260, loss: 3.23963540722616e-05
step: 270, loss: 0.0003655760665424168
```

```
step: 280, loss: 3.439020656514913e-05
step: 290, loss: 3.2084051781566814e-05
step: 300, loss: 9.137437155004591e-05
step: 310, loss: 0.0005630278028547764
step: 320, loss: 0.00046480362652800977
step: 330, loss: 3.404254312044941e-05
step: 340, loss: 0.0033134575933218002
step: 350, loss: 4.294705286156386e-05
step: 360, loss: 0.002352513140067458
step: 370, loss: 3.708307849592529e-05
step: 380, loss: 3.579805706976913e-05
step: 390, loss: 0.00013728890917263925
step: 400, loss: 3.6152603570371866e-05
step: 410, loss: 4.465788515517488e-05
step: 420, loss: 7.830376853235066e-05
step: 430, loss: 3.676023698062636e-05
step: 440, loss: 2.6686209821491502e-05
step: 450, loss: 4.3359934352338314e-05
step: 460, loss: 0.0002286985400132835
step: 470, loss: 3.348126483615488e-05
step: 480, loss: 0.00010068802657769993
step: 490, loss: 3.405056850169785e-05
step: 500, loss: 0.00011326653475407511
step: 510, loss: 7.576767529826611e-05
step: 520, loss: 3.0583883926738054e-05
step: 530, loss: 4.461973730940372e-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=0.9031365313653137, f1=0.9040590405904059,
best_f1=0.8996350364963502
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 2.9082353648846038e-05
step: 10, loss: 0.00014682256733067334
step: 20, loss: 0.0022541021462529898
step: 30, loss: 4.536029155133292e-05
step: 40, loss: 4.4874097511637956e-05
step: 50, loss: 5.5476197303505614e-05
step: 60, loss: 2.6898787837126292e-05
step: 70, loss: 0.0005296061863191426
step: 80, loss: 0.00027701156795956194
step: 90, loss: 2.2581998564419337e-05
step: 100, loss: 2.805427720886655e-05
step: 110, loss: 0.010629831813275814
step: 120, loss: 0.00414556544274917
step: 130, loss: 4.165929931332357e-05
```

```
step: 140, loss: 1.9699038602993824e-05
step: 150, loss: 0.000846135662868619
step: 160, loss: 3.425617978791706e-05
step: 170, loss: 0.0015625106170773506
step: 180, loss: 1.9468005120870657e-05
step: 190, loss: 0.0001456224563298747
step: 200, loss: 3.0054552553338e-05
step: 210, loss: 0.0003772648051381111
step: 220, loss: 0.00018872346845455468
step: 230, loss: 4.0230843296740204e-05
step: 240, loss: 4.854729195358232e-05
step: 250, loss: 0.00012933477410115302
step: 260, loss: 7.432371057802811e-05
step: 270, loss: 3.1779531127540395e-05
step: 280, loss: 2.9943017580080777e-05
step: 290, loss: 0.0016894302098080516
step: 300, loss: 0.002648429712280631
step: 310, loss: 0.0005507060559466481
step: 320, loss: 0.0002338803024031222
step: 330, loss: 0.0035418409388512373
step: 340, loss: 0.000346727785654366
step: 350, loss: 5.7929788454202935e-05
step: 360, loss: 0.00010648761235643178
step: 370, loss: 1.9885268557118252e-05
step: 380, loss: 0.00010674069199012592
step: 390, loss: 5.5171312851598486e-05
step: 400, loss: 0.00023212005908135325
step: 410, loss: 7.340782758546993e-05
step: 420, loss: 4.5631368266185746e-05
step: 430, loss: 2.410958768450655e-05
step: 440, loss: 0.0001722830202197656
step: 450, loss: 8.100333798211068e-05
step: 460, loss: 2.2473914214060642e-05
step: 470, loss: 4.337686186772771e-05
step: 480, loss: 0.00017660630692262203
step: 490, loss: 3.2658288546372205e-05
step: 500, loss: 0.0009072293760254979
step: 510, loss: 0.0006938075530342758
step: 520, loss: 4.457091199583374e-05
step: 530, loss: 2.803193092404399e-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=0.897423887587822, f1=0.8989287377736376,
best_f1=0.8996350364963502
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.")
5742it [00:22, 255.75it/s]
load_f1 = 0.9053613053613055
real_f1 = 0.90181479758027
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 [00:16, 265.70it/s]
```

2.1.4 Amazon-Google - Running the matcher

```python
# 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 \
```

174

```
  --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   : 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: 0.8337807655334473
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.06548142433166504
step: 20, loss: 0.38525962829589844
step: 30, loss: 0.374788373708725
step: 40, loss: 0.5033724904060364
step: 50, loss: 0.3111454248428345
step: 60, loss: 0.3680569529533386
step: 70, loss: 0.2767786681652069
step: 80, loss: 0.2903327941894531
step: 90, loss: 0.43282729387283325
step: 100, loss: 0.2011963278055191
step: 110, loss: 0.29559141397476196
step: 120, loss: 0.25695788860321045
step: 130, loss: 0.23154965043067932
step: 140, loss: 0.31882378458976746
step: 150, loss: 0.3208945691585541
step: 160, loss: 0.26309022307395935
step: 170, loss: 0.19286341965198517
step: 180, loss: 0.20009249448776245
step: 190, loss: 0.27198657393455505
step: 200, loss: 0.1905180662870407
step: 210, loss: 0.5201736092567444
/usr/local/lib/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=0.3961038961038961, f1=0.39781021897810215,
best_f1=0.39781021897810215
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08718249946832657
step: 10, loss: 0.10454125702381134
step: 20, loss: 0.2842254340648651
step: 30, loss: 0.19628357887268066
step: 40, loss: 0.06034741923213005
```

```
step: 50, loss: 0.22324667870998383
step: 60, loss: 0.1572829633951187
step: 70, loss: 0.288668692111969
step: 80, loss: 0.2895306646823883
step: 90, loss: 0.1382617950439453
step: 100, loss: 0.18277132511138916
step: 110, loss: 0.04430116340517998
step: 120, loss: 0.2696143388748169
step: 130, loss: 0.2287580668926239
step: 140, loss: 0.3099009394645691
step: 150, loss: 0.25548022985458374
step: 160, loss: 0.1698032170534134
step: 170, loss: 0.2592656910419464
step: 180, loss: 0.37527740001678467
step: 190, loss: 0.1874864548444748
step: 200, loss: 0.2811562418937683
step: 210, loss: 0.3726328909397125
/usr/local/lib/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=0.4609800362976406, f1=0.5018315018315018,
best_f1=0.5018315018315018
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1875815987586975
step: 10, loss: 0.24759399890899658
step: 20, loss: 0.3024594485759735
step: 30, loss: 0.07400152087211609
step: 40, loss: 0.18351198732852936
step: 50, loss: 0.21474589407444
step: 60, loss: 0.3609232008457184
step: 70, loss: 0.19411729276180267
step: 80, loss: 0.14800406992435455
step: 90, loss: 0.16429336369037628
step: 100, loss: 0.1340516358613968
step: 110, loss: 0.2770861089229584
step: 120, loss: 0.206455796957016
step: 130, loss: 0.2239028513431549
step: 140, loss: 0.3639140725135803
step: 150, loss: 0.25400272011756897
step: 160, loss: 0.10782173275947571
step: 170, loss: 0.16466505825519562
step: 180, loss: 0.09980249404907227
```

```
step: 190, loss: 0.2582845389842987
step: 200, loss: 0.2198481261730194
step: 210, loss: 0.25736141204833984
/usr/local/lib/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=0.4752475247524752, f1=0.516, best_f1=0.516
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2922063171863556
step: 10, loss: 0.06015527993440628
step: 20, loss: 0.14262156188488007
step: 30, loss: 0.07847729325294495
step: 40, loss: 0.15821778774261475
step: 50, loss: 0.15349358320236206
step: 60, loss: 0.0814436599612236
step: 70, loss: 0.16483674943447113
step: 80, loss: 0.1837759166955948
step: 90, loss: 0.06196345388889313
step: 100, loss: 0.17996154725551605
step: 110, loss: 0.11801635473966599
step: 120, loss: 0.07300838083028793
step: 130, loss: 0.235294371843338
step: 140, loss: 0.10578346997499466
step: 150, loss: 0.21834829449653625
step: 160, loss: 0.11539751291275024
step: 170, loss: 0.13497033715248108
step: 180, loss: 0.15238328278064728
step: 190, loss: 0.1744106560945511
step: 200, loss: 0.2379525750875473
step: 210, loss: 0.04973411187529564
/usr/local/lib/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=0.4861407249466951, f1=0.4708333333333334,
best_f1=0.4708333333333334
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.05742418020963669
step: 10, loss: 0.0904601588845253
step: 20, loss: 0.04822332039475441
step: 30, loss: 0.03771226853132248
step: 40, loss: 0.06893022358417511
step: 50, loss: 0.1665831357240677
step: 60, loss: 0.028278274461627007
step: 70, loss: 0.04926125332713127
step: 80, loss: 0.08993957191705704
step: 90, loss: 0.0323953814804554
step: 100, loss: 0.14056800305843353
step: 110, loss: 0.02262728475034237
step: 120, loss: 0.037716422230005264
step: 130, loss: 0.03709957003593445
step: 140, loss: 0.054537706077098846
step: 150, loss: 0.03805225342512131
step: 160, loss: 0.0316318273544311
step: 170, loss: 0.1303476244211197
step: 180, loss: 0.2674318552017212
step: 190, loss: 0.06410899013280869
step: 200, loss: 0.08993306010961533
step: 210, loss: 0.12538883090019226
/usr/local/lib/python3.7/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=0.456140350877193, f1=0.4657933042212518,
best_f1=0.4708333333333334
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05813221633434296
step: 10, loss: 0.014986395835876465
step: 20, loss: 0.020977944135665894
step: 30, loss: 0.07489190995693207
step: 40, loss: 0.09706646203994751
step: 50, loss: 0.0597965233027935
step: 60, loss: 0.15155312418937683
step: 70, loss: 0.003715973813086748
step: 80, loss: 0.2203749269247055
step: 90, loss: 0.201389878988266
step: 100, loss: 0.0378678143024446
step: 110, loss: 0.03243817016482353
step: 120, loss: 0.03918377682566643
step: 130, loss: 0.008373919874429703
step: 140, loss: 0.1493033468723297
step: 150, loss: 0.06867644190788269
step: 160, loss: 0.19728641211986542
step: 170, loss: 0.09200804680585861
```

```
step: 180, loss: 0.029924873262643814
step: 190, loss: 0.036646779626607895
step: 200, loss: 0.019749047234654427
step: 210, loss: 0.03712419420480728
/usr/local/lib/python3.7/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=0.4444444444444445, f1=0.463768115942029,
best_f1=0.4708333333333334
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09640754014253616
step: 10, loss: 0.007864260114729404
step: 20, loss: 0.07015971839427948
step: 30, loss: 0.020235704258084297
step: 40, loss: 0.011621496640145779
step: 50, loss: 0.13357461988925934
step: 60, loss: 0.16390188038349152
step: 70, loss: 0.05237138271331787
step: 80, loss: 0.014513658359646797
step: 90, loss: 0.10776166617870331
step: 100, loss: 0.04401580989360809
step: 110, loss: 0.0329226590692997
step: 120, loss: 0.05745599791407585
step: 130, loss: 0.00554531579837203
step: 140, loss: 0.0099534261971171211
step: 150, loss: 0.03129784017801285
step: 160, loss: 0.0706164687871933
step: 170, loss: 0.007633023429661989
step: 180, loss: 0.0018164686625823379
step: 190, loss: 0.07399636507034302
step: 200, loss: 0.03398749232292175
step: 210, loss: 0.13271376490592957
/usr/local/lib/python3.7/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=0.4196277495769882, f1=0.4664429530201342,
best_f1=0.4708333333333334
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.032599687576293945
step: 10, loss: 0.009068922139704227
step: 20, loss: 0.11694194376468658
step: 30, loss: 0.00202498328872025
step: 40, loss: 0.22008909285068512
step: 50, loss: 0.02966858632862568
```

```
step: 60, loss: 0.07434888184070587
step: 70, loss: 0.018977778032422066
step: 80, loss: 0.07624726742506027
step: 90, loss: 0.1176474541425705
step: 100, loss: 0.00290561455488205
step: 110, loss: 0.0016458820318803191
step: 120, loss: 0.03270075470209122
step: 130, loss: 0.007593114860355854
step: 140, loss: 0.07242237776517868
step: 150, loss: 0.004248755984008312
step: 160, loss: 0.08771200478076935
step: 170, loss: 0.0037405493203550577
step: 180, loss: 0.042301155626773834
step: 190, loss: 0.023117559030652046
step: 200, loss: 0.09406159073114395
step: 210, loss: 0.10915740579366684
/usr/local/lib/python3.7/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=0.4051724137931035, f1=0.440251572327044,
best_f1=0.4708333333333334
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10958531498908997
step: 10, loss: 0.007302991580218077
step: 20, loss: 0.12080637365579605
step: 30, loss: 0.019734680652618408
step: 40, loss: 0.015672490000724792
step: 50, loss: 0.0017717719310894608
step: 60, loss: 0.022497154772281647
step: 70, loss: 0.0010308617493137717
step: 80, loss: 0.0023329509422183037
step: 90, loss: 0.15136368572711945
step: 100, loss: 0.010130520910024643
step: 110, loss: 0.005727812182158232
step: 120, loss: 0.0030139987356960773
step: 130, loss: 0.004736994858831167
step: 140, loss: 0.04507861286401749
step: 150, loss: 0.008736250922083855
step: 160, loss: 0.0005062891286797822
step: 170, loss: 0.002866193884983659
step: 180, loss: 0.03321181237697601
step: 190, loss: 0.03629881143569946
step: 200, loss: 0.04581199213862419
step: 210, loss: 0.0091177923604846
/usr/local/lib/python3.7/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=0.39565217391304347, f1=0.4312896405919662,
best_f1=0.4708333333333334
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0035030031576752663
step: 10, loss: 0.0013775225961580873
step: 20, loss: 0.06852558255195618
step: 30, loss: 0.008012042380869389
step: 40, loss: 0.005351515021175146
step: 50, loss: 0.028457464650273323
step: 60, loss: 0.029855862259864807
step: 70, loss: 0.010648488998413086
step: 80, loss: 0.00038763193879276514
step: 90, loss: 0.01760866492986679
step: 100, loss: 0.00036028659087605774
step: 110, loss: 0.0015128821833059192
step: 120, loss: 0.02239670604467392
step: 130, loss: 0.03578922152519226
step: 140, loss: 0.00030999939190223813
step: 150, loss: 0.07170045375823975
step: 160, loss: 0.02475845441222191
step: 170, loss: 0.002597522223368287
step: 180, loss: 0.003173135919496417
step: 190, loss: 0.051594484597444534
step: 200, loss: 0.003044651821255684
step: 210, loss: 0.14183150231838226
/usr/local/lib/python3.7/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=0.4146868250539957, f1=0.4273858921161826,
best_f1=0.4708333333333334
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003037302754819393
step: 10, loss: 0.03149773180484772
step: 20, loss: 0.002392767695710063
step: 30, loss: 0.006444964557886124
step: 40, loss: 0.02964616008102894
step: 50, loss: 0.002500551287084818
step: 60, loss: 0.0077479928731918335
step: 70, loss: 0.006069390568882227
step: 80, loss: 0.08718675374984741
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 90, loss: 0.0005559442797675729
step: 100, loss: 0.0007986897835507989
```

```
step: 110, loss: 0.0014176114927977324
step: 120, loss: 0.003062909236177802
step: 130, loss: 0.02170775644481182
step: 140, loss: 0.0025500573683530092
step: 150, loss: 0.05591044947504997
step: 160, loss: 0.005129865370690823
step: 170, loss: 0.08491266518831253
step: 180, loss: 0.014971202239394188
step: 190, loss: 0.011914132162928581
step: 200, loss: 0.00019417982548475266
step: 210, loss: 0.00015084139886312187
/usr/local/lib/python3.7/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=0.3963553530751709, f1=0.42035398230088494,
best_f1=0.4708333333333334
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003051003150176257
step: 10, loss: 0.00023666716879233718
step: 20, loss: 0.030331384390592575
step: 30, loss: 0.0026930964086204767
step: 40, loss: 0.0018355417996644974
step: 50, loss: 0.005720323882997036
step: 60, loss: 0.0010507210390642285
step: 70, loss: 0.014923356473445892
step: 80, loss: 0.0035776584409177303
step: 90, loss: 0.048487935215234756
step: 100, loss: 0.0038705940824002028
step: 110, loss: 0.00038271918310783803
step: 120, loss: 0.007163667120039463
step: 130, loss: 0.021083394065499306
step: 140, loss: 0.001294023240916431
step: 150, loss: 0.00028478316380642354
step: 160, loss: 0.004532424733042717
step: 170, loss: 0.0006357172969728708
step: 180, loss: 0.0014851113082841039
step: 190, loss: 0.024980653077363968
step: 200, loss: 0.005054258741438389
step: 210, loss: 0.04165303707122803
/usr/local/lib/python3.7/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=0.41282565130260523, f1=0.45436893203883494,
best_f1=0.4708333333333334
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00018629670375958085
step: 10, loss: 0.02149646356701851
step: 20, loss: 0.0010000875918194652
step: 30, loss: 0.002572194905951619
step: 40, loss: 0.0005530782509595156
step: 50, loss: 0.0007320474833250046
step: 60, loss: 0.008802159689366817
step: 70, loss: 0.05704515054821968
step: 80, loss: 0.018053123727440834
step: 90, loss: 0.004520464688539505
step: 100, loss: 0.0025667890440672636
step: 110, loss: 0.000768783560488373
step: 120, loss: 0.00034222594695165753
step: 130, loss: 0.000482038565678522
step: 140, loss: 0.0008131967624649405
step: 150, loss: 0.005545966327190399
step: 160, loss: 0.007298201322555542
step: 170, loss: 0.00191679154522717
step: 180, loss: 0.0026509903836995363
step: 190, loss: 0.08991855382919312
step: 200, loss: 0.0006213897140696645
step: 210, loss: 0.0006260970840230584
/usr/local/lib/python3.7/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=0.4036281179138322, f1=0.4370860927152318,
best_f1=0.4708333333333334
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005784186068922281
step: 10, loss: 0.0005686279619112611
step: 20, loss: 0.00021497112174984068
step: 30, loss: 0.010891047306358814
step: 40, loss: 0.005044637247920036
step: 50, loss: 0.0006287121796049178
step: 60, loss: 0.0004871422133874148
step: 70, loss: 0.0018616531742736697
step: 80, loss: 0.0038368008099496365
step: 90, loss: 0.012557776644825935
step: 100, loss: 0.0023448707070201635
step: 110, loss: 0.0006142161437310278
step: 120, loss: 0.0019989751745015383
step: 130, loss: 0.0009262601379305124
step: 140, loss: 0.0060825832188129425
step: 150, loss: 0.03354710713028908
step: 160, loss: 0.0010401932522654533
```

```
step: 170, loss: 0.004052888602018356
step: 180, loss: 0.00017033389303833246
step: 190, loss: 0.0015600415645167232
step: 200, loss: 0.0031501834746450186
step: 210, loss: 0.00019078998593613505
/usr/local/lib/python3.7/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=0.4053452115812918, f1=0.4292035398230089,
best_f1=0.4708333333333334
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012004951946437359
step: 10, loss: 0.0028956751339137554
step: 20, loss: 0.0007842733175493777
step: 30, loss: 0.002110114088281989
step: 40, loss: 0.000365749147022143
step: 50, loss: 0.00017067790031433105
step: 60, loss: 0.0008544449810869992
step: 70, loss: 0.0011217038845643401
step: 80, loss: 0.002320552011951804
step: 90, loss: 0.0005601731245405972
step: 100, loss: 0.0004126471758354455
step: 110, loss: 0.00010002106864703819
step: 120, loss: 0.001671033096499741
step: 130, loss: 0.0005591695080511272
step: 140, loss: 0.00171552668325603
step: 150, loss: 0.001859582494944334
step: 160, loss: 0.00029184570303186774
step: 170, loss: 0.0002890037721954286
step: 180, loss: 0.01278870739042759
step: 190, loss: 0.026795770972967148
step: 200, loss: 0.005713484715670347
step: 210, loss: 0.0016718414844945073
/usr/local/lib/python3.7/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=0.402745995423341, f1=0.43595505617977526,
best_f1=0.4708333333333334
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.")
2293it [00:06, 329.11it/s]
load_f1 = 0.47844827586206895
real_f1 = 0.4861407249466951
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 [00:16, 261.04it/s]
```

2.1.5 Walmart-Amazon - Running the matcher

```python
# 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 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: 0.8761487603187561
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.16827009618282318
step: 20, loss: 0.1592557281255722
step: 30, loss: 0.50071257352828988
step: 40, loss: 0.25315526127815247
step: 50, loss: 0.30820798873901367
step: 60, loss: 0.36519694328308105
step: 70, loss: 0.17030636966228485
step: 80, loss: 0.5063663125038147
step: 90, loss: 0.23614449799060822
step: 100, loss: 0.22451668977737427
step: 110, loss: 0.2368757128715515
step: 120, loss: 0.41990941762924194
step: 130, loss: 0.3385409712791443
step: 140, loss: 0.3388080298900604
step: 150, loss: 0.2692643702030182
step: 160, loss: 0.23157714307308197
step: 170, loss: 0.3980990946292877
step: 180, loss: 0.30871450901031494
step: 190, loss: 0.15105821192264557
/usr/local/lib/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=0.4314606741573034, f1=0.41986455981941306,
best_f1=0.41986455981941306
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.25304341316223145
step: 10, loss: 0.04738935828208923
step: 20, loss: 0.14744414389133453
step: 30, loss: 0.17156334221363068
step: 40, loss: 0.47651782631874084
step: 50, loss: 0.2933870851993561
step: 60, loss: 0.1367240846157074
step: 70, loss: 0.23506928980350494
step: 80, loss: 0.20613205432891846
step: 90, loss: 0.10064970701932907
step: 100, loss: 0.3008871376514435
step: 110, loss: 0.19395211338996887
```

```
step: 120, loss: 0.3134462237358093
step: 130, loss: 0.19078773260116577
step: 140, loss: 0.3469194173812866
step: 150, loss: 0.04311797767877579
step: 160, loss: 0.0805833712220192
step: 170, loss: 0.21672265231609344
step: 180, loss: 0.13882797956466675
step: 190, loss: 0.2622933089733124
/usr/local/lib/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=0.706896551724138, f1=0.7228260869565217,
best_f1=0.7228260869565217
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.21447128057479858
step: 10, loss: 0.244327113032341
step: 20, loss: 0.024613751098513603
step: 30, loss: 0.1996963918209076
step: 40, loss: 0.06178075075149536
step: 50, loss: 0.055451489984989166
step: 60, loss: 0.1260284036397934
step: 70, loss: 0.20791411399841309
step: 80, loss: 0.1738048791885376
step: 90, loss: 0.19979935884475708
step: 100, loss: 0.07958279550075531
step: 110, loss: 0.27655118703842163
step: 120, loss: 0.12106871604919434
step: 130, loss: 0.02138308435678482
step: 140, loss: 0.19427905976772308
step: 150, loss: 0.2801416516304016
step: 160, loss: 0.12116070091724396
step: 170, loss: 0.05712412670254707
step: 180, loss: 0.17120042443275452
step: 190, loss: 0.21003364026546478
/usr/local/lib/python3.7/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=0.6900269541778975, f1=0.6881720430107526,
best_f1=0.7228260869565217
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.10311809927225113
step: 10, loss: 0.05733533948659897
step: 20, loss: 0.05356263741850853
step: 30, loss: 0.046380504965782166
step: 40, loss: 0.005314954556524754
step: 50, loss: 0.04616566374897957
step: 60, loss: 0.19066070020198822
step: 70, loss: 0.04183505102992058
step: 80, loss: 0.04199988767504692
step: 90, loss: 0.038278013467788696
step: 100, loss: 0.03942789137363434
step: 110, loss: 0.055064745247364044
step: 120, loss: 0.09546563029289246
step: 130, loss: 0.39298707246780396
step: 140, loss: 0.12463338673114777
step: 150, loss: 0.017651595175266266
step: 160, loss: 0.03338024392724037
step: 170, loss: 0.020626427605748177
step: 180, loss: 0.20611952245235443
step: 190, loss: 0.03187886252999306
/usr/local/lib/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=0.7401129943502824, f1=0.7520891364902507,
best_f1=0.7520891364902507
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02681226097047329
step: 10, loss: 0.020974621176719666
step: 20, loss: 0.0243227481842041
step: 30, loss: 0.005657027941197157
step: 40, loss: 0.015185238793492317
step: 50, loss: 0.08472563326358795
step: 60, loss: 0.05135254189372063
step: 70, loss: 0.014097688719630241
step: 80, loss: 0.004956577904522419
step: 90, loss: 0.0013631567126139998
step: 100, loss: 0.020815202966332436
step: 110, loss: 0.007012709975242615
step: 120, loss: 0.0034050203394144773
step: 130, loss: 0.04882100969552994
step: 140, loss: 0.007921307347714901
step: 150, loss: 0.0123963113874197
```

```
step: 160, loss: 0.0026515612844377756
step: 170, loss: 0.009996980428695679
step: 180, loss: 0.006493085995316505
step: 190, loss: 0.13411729037761688
/usr/local/lib/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=0.7666666666666667, f1=0.7465181058495822,
best_f1=0.7465181058495822
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.024876780807971954
step: 10, loss: 0.0018207598477602005
step: 20, loss: 0.002617559162899852
step: 30, loss: 0.0025219458620995283
step: 40, loss: 0.04395414516329765
step: 50, loss: 0.01038337592035532
step: 60, loss: 0.0019248644821345806
step: 70, loss: 0.05368699133396149
step: 80, loss: 0.16727857291698456
step: 90, loss: 0.030846361070871353
step: 100, loss: 0.058376286178827286
step: 110, loss: 0.018717506900429726
step: 120, loss: 0.01555632520467043
step: 130, loss: 0.04343317076563835
step: 140, loss: 0.01992988772690296
step: 150, loss: 0.043862033635377884
step: 160, loss: 0.016390802338719368
step: 170, loss: 0.1462383270263672
step: 180, loss: 0.09499972313642502
step: 190, loss: 0.0346779003739357
/usr/local/lib/python3.7/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=0.7262872628726287, f1=0.7533875338753386,
best_f1=0.7465181058495822
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0017388706328347325
step: 10, loss: 0.10212720185518265
step: 20, loss: 0.0007435539155267179
step: 30, loss: 0.09567087143659592
```

```
step: 40, loss: 0.07744546979665756
step: 50, loss: 0.03422486037015915
step: 60, loss: 0.03879014030098915
step: 70, loss: 0.0011421616654843092
step: 80, loss: 0.0098071424290553783
step: 90, loss: 0.04443015530705452
step: 100, loss: 0.002154221758246422
step: 110, loss: 0.019095011055469513
step: 120, loss: 0.005684332922101021
step: 130, loss: 0.029797369614243507
step: 140, loss: 0.0512615330517292
step: 150, loss: 0.0030809720046818256
step: 160, loss: 0.0011483338894322515
step: 170, loss: 0.009227112866938114
step: 180, loss: 0.0023341563064604998
step: 190, loss: 0.008419702760875225
/usr/local/lib/python3.7/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=0.7341772151898734, f1=0.741687979539642,
best_f1=0.7465181058495822
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005660993629135191
step: 10, loss: 0.0013277410762384534
step: 20, loss: 0.0005389744183048606
step: 30, loss: 0.0032787343952804804
step: 40, loss: 0.001134060206823051
step: 50, loss: 0.0033166310749948025
step: 60, loss: 0.0022665332071483135
step: 70, loss: 0.0016868062084540725
step: 80, loss: 0.004773894790560007
step: 90, loss: 0.0007769943331368268
step: 100, loss: 0.0028301943093538284
step: 110, loss: 0.0006834767409600317
step: 120, loss: 0.0006467464845627546
step: 130, loss: 0.1663864552974701
step: 140, loss: 0.001501482562161982
step: 150, loss: 0.003353218315169215
step: 160, loss: 0.0049040955491364
step: 170, loss: 0.010556455701589584
step: 180, loss: 0.002685681451112032
step: 190, loss: 0.0027220153715461493
/usr/local/lib/python3.7/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=0.7365853658536586, f1=0.7146401985111663,
```

```
best_f1=0.7465181058495822
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.026014087721705437
step: 10, loss: 0.004884480964392424
step: 20, loss: 0.029052842408418655
step: 30, loss: 0.0025337617844343185
step: 40, loss: 0.0038584235589951277
step: 50, loss: 0.004719422664493322
step: 60, loss: 0.0003912674728780985
step: 70, loss: 0.0013381425524130464
step: 80, loss: 0.0009906056802719831
step: 90, loss: 0.04980267956852913
step: 100, loss: 0.002725412603467703
step: 110, loss: 0.15595610439777374
step: 120, loss: 0.07787436246871948
step: 130, loss: 0.005393470171838999
step: 140, loss: 0.0005075006047263741
step: 150, loss: 0.06326925754547119
step: 160, loss: 0.0015701957745477557
step: 170, loss: 0.003272648900747299
step: 180, loss: 0.0023901392705738544
step: 190, loss: 0.0021480161231011152
/usr/local/lib/python3.7/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=0.7379134860050889, f1=0.7185929648241205,
best_f1=0.7465181058495822
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003055390843655914
step: 10, loss: 0.02103791944682598
step: 20, loss: 0.0025192287284880877
step: 30, loss: 0.003224650863558054
step: 40, loss: 0.0044760447926819324
step: 50, loss: 0.0003558742464520037
step: 60, loss: 0.12085886299610138
step: 70, loss: 0.0009713189792819321
step: 80, loss: 0.0004306879418436438
step: 90, loss: 0.000607224996201694
step: 100, loss: 0.0049642049707472324
step: 110, loss: 0.0020091868937015533
step: 120, loss: 0.004663917701691389
step: 130, loss: 0.0015213579172226851
step: 140, loss: 0.04392214119434357
step: 150, loss: 0.00026236657868139446
```

```
step: 160, loss: 0.0002346217370359227
step: 170, loss: 0.00030735658947378397
step: 180, loss: 0.0011574814561754465
step: 190, loss: 0.022845566272735596
/usr/local/lib/python3.7/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=0.7374005305039788, f1=0.7180851063829788,
best_f1=0.7465181058495822
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001191004877910018
step: 10, loss: 0.035642992705106735
step: 20, loss: 0.00120980304200020223
step: 30, loss: 0.007105839904397726
step: 40, loss: 0.0017981813289225101
step: 50, loss: 0.0008245132048614323
step: 60, loss: 0.053585492074489594
step: 70, loss: 0.012769915163516998
step: 80, loss: 0.0001479166530771181
step: 90, loss: 0.00786976981908083
step: 100, loss: 0.003235525218769908
step: 110, loss: 0.03563076630234718
step: 120, loss: 0.00045039504766464233
step: 130, loss: 0.00037298694951459765
step: 140, loss: 0.00021453676163218915
step: 150, loss: 0.0033998102881014347
step: 160, loss: 0.0001982936664717272
step: 170, loss: 0.0005462515400722623
step: 180, loss: 0.025367528200149536
step: 190, loss: 0.000774907588493079
/usr/local/lib/python3.7/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=0.7422680412371134, f1=0.7480916030534353,
best_f1=0.7465181058495822
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002339248778298497
step: 10, loss: 0.0009859782876446843
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 20, loss: 0.0004727219056803733
step: 30, loss: 0.0009032501839101315
step: 40, loss: 0.0003196817706339061
step: 50, loss: 0.013870255090296268
step: 60, loss: 0.00016929904813878238
```

```
step: 70, loss: 0.0002451046311762184
step: 80, loss: 0.0006968540255911648
step: 90, loss: 0.0003556654555723071
step: 100, loss: 0.0003302493423689157
step: 110, loss: 0.0036323636304587126
step: 120, loss: 0.00022856007853988558
step: 130, loss: 0.0001270551438210532
step: 140, loss: 0.0004376911383587867
step: 150, loss: 0.0009217443875968456
step: 160, loss: 0.0007663911674171686
step: 170, loss: 0.014804001897573471
step: 180, loss: 0.001679292763583362
step: 190, loss: 0.015818068757653236
/usr/local/lib/python3.7/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=0.7231920199501246, f1=0.7002518891687657,
best_f1=0.7465181058495822
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00039677994209341705
step: 10, loss: 0.00018699257634580135
step: 20, loss: 0.0005322537035681307
step: 30, loss: 0.00026210182113572955
step: 40, loss: 0.00013204799324739724
step: 50, loss: 0.0002107013133354485
step: 60, loss: 0.018055498600006104
step: 70, loss: 0.009160938672721386
step: 80, loss: 0.00024024331651162356
step: 90, loss: 0.003010894637554884
step: 100, loss: 0.00020403796224854887
step: 110, loss: 0.00019991249428130686
step: 120, loss: 0.000559369451366365
step: 130, loss: 0.00042651151306927204
step: 140, loss: 0.005035865120589733
step: 150, loss: 0.00011000613449141383
step: 160, loss: 0.0007133673643693328
step: 170, loss: 0.00018262346566188908
step: 180, loss: 0.00041425428935326636
step: 190, loss: 0.0003543277271091938
/usr/local/lib/python3.7/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=0.7291666666666667, f1=0.7150537634408601,
best_f1=0.7465181058495822
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004072831361554563
step: 10, loss: 0.0035179746337234974
step: 20, loss: 0.0008117855759337544
step: 30, loss: 0.0005686834920197725
step: 40, loss: 0.0007530883885920048
step: 50, loss: 0.015476394444704056
step: 60, loss: 0.0001446992246201262
step: 70, loss: 0.00013066575047560036
step: 80, loss: 8.046428411034867e-05
step: 90, loss: 0.0002372709714109078
step: 100, loss: 0.00024151831166818738
step: 110, loss: 0.00014337255561258644
step: 120, loss: 0.00011191137309651822
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 130, loss: 0.0001232007925864309
step: 140, loss: 0.0003091042162850499
step: 150, loss: 0.00017258110165130347
step: 160, loss: 0.00013163255061954266
step: 170, loss: 0.00018546245701145388
step: 180, loss: 0.00017939710232894868
step: 190, loss: 0.0006391702336259186
/usr/local/lib/python3.7/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=0.7286432160804021, f1=0.7070707070707071,
best_f1=0.7465181058495822
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00040546301170252264
step: 10, loss: 9.235880133928731e-05
step: 20, loss: 0.00023153467918746173
step: 30, loss: 0.00014647688658442348
step: 40, loss: 0.010306903161108494
step: 50, loss: 0.0002502302813809365
step: 60, loss: 0.0001278312411159277
step: 70, loss: 0.001043430995196104
step: 80, loss: 0.0011130711063742638
step: 90, loss: 0.012854116968810558
step: 100, loss: 0.00011085846199421212
step: 110, loss: 0.0007992989849299192
step: 120, loss: 0.05724530667066574
step: 130, loss: 0.0001366051146760583
step: 140, loss: 0.00033742416417226195
step: 150, loss: 0.0019611159805208445
step: 160, loss: 0.0003101261972915381
step: 170, loss: 0.00011800304491771385
```

```
step: 180, loss: 0.0004358137084636837
step: 190, loss: 0.0016532279551029205
/usr/local/lib/python3.7/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=0.7295285359801489, f1=0.7061728395061728,
best_f1=0.7465181058495822
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.")
2049it [00:08, 228.22it/s]
load_f1 = 0.7611111111111111
real_f1 = 0.7632311977715878
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 [00:16, 260.57it/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 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: 0.8608489632606506
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.2306361198425293
step: 20, loss: 0.15786631405353546
step: 30, loss: 0.22978875041007996
step: 40, loss: 0.30703914165496826
step: 50, loss: 0.37227678298950195
step: 60, loss: 0.43672850728034973
step: 70, loss: 0.30695319175720215
step: 80, loss: 0.26054614782333374
step: 90, loss: 0.4046425223350525
step: 100, loss: 0.2357613742351532
step: 110, loss: 0.1955546736717224
step: 120, loss: 0.5353491902351379
step: 130, loss: 0.42340344190597534
step: 140, loss: 0.500351071357727
step: 150, loss: 0.10675344616174698
step: 160, loss: 0.33875319361686707
step: 170, loss: 0.24890460073947906
/usr/local/lib/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=0.3233695652173913, f1=0.2845070422535212,

```
best_f1=0.2845070422535212
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.4072779715061188
step: 10, loss: 0.17043247818946838
step: 20, loss: 0.3252335786819458
step: 30, loss: 0.10441413521766663
step: 40, loss: 0.2601054310798645
step: 50, loss: 0.2427932769060135
step: 60, loss: 0.18038034439086914
step: 70, loss: 0.33970171213150024
step: 80, loss: 0.13977079093456268
step: 90, loss: 0.18667010962963104
step: 100, loss: 0.1401086449623108
step: 110, loss: 0.2710365056991577
step: 120, loss: 0.19576366245746613
step: 130, loss: 0.09868723154067993
step: 140, loss: 0.18892794847488403
step: 150, loss: 0.14060582220554352
step: 160, loss: 0.13928410410881042
step: 170, loss: 0.11403030157089233
/usr/local/lib/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=0.6578366445916115, f1=0.6285714285714287,
best_f1=0.6285714285714287
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.15102450549602509
step: 10, loss: 0.06257514655590057
step: 20, loss: 0.042643673717975616
step: 30, loss: 0.15827889740467072
step: 40, loss: 0.0784616619348526
step: 50, loss: 0.13982756435871124
step: 60, loss: 0.3209995925426483
step: 70, loss: 0.09058491140604019
step: 80, loss: 0.10672695934772491
step: 90, loss: 0.06594245135784149
step: 100, loss: 0.06691823899745941
step: 110, loss: 0.05932944640517235
step: 120, loss: 0.025978952646255493
step: 130, loss: 0.1581786870956421
```

```
step: 140, loss: 0.010054334066808224
step: 150, loss: 0.10858942568302155
step: 160, loss: 0.04883319139480591
step: 170, loss: 0.0863766223192215
/usr/local/lib/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=0.7378190255220417, f1=0.6859688195991092,
best_f1=0.6859688195991092
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06646993011236191
step: 10, loss: 0.22583507001399994
step: 20, loss: 0.054685723036527634
step: 30, loss: 0.09091121703386307
step: 40, loss: 0.11596576124429703
step: 50, loss: 0.01485427562147379
step: 60, loss: 0.10364589840173721
step: 70, loss: 0.008422932587563992
step: 80, loss: 0.10685736685991287
step: 90, loss: 0.028982147574424744
step: 100, loss: 0.01848428137600422
step: 110, loss: 0.06877042353153229
step: 120, loss: 0.08383633196353912
step: 130, loss: 0.024290766566991806
step: 140, loss: 0.010420718230307102
step: 150, loss: 0.04859531670808792
step: 160, loss: 0.062092412263154984
step: 170, loss: 0.0447167232632637
/usr/local/lib/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=0.7428571428571429, f1=0.7135922330097088,
best_f1=0.7135922330097088
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.025563135743141174
step: 10, loss: 0.03206585347652435
```

```
step: 20, loss: 0.05359407514333725
step: 30, loss: 0.047410257160663605
step: 40, loss: 0.01844029314815998
step: 50, loss: 0.026204640045762062
step: 60, loss: 0.00432957848533988
step: 70, loss: 0.06607375293970108
step: 80, loss: 0.005234838929027319
step: 90, loss: 0.01804163306951523
step: 100, loss: 0.039255235344171524
step: 110, loss: 0.010012791492044926
step: 120, loss: 0.18006351590156555
step: 130, loss: 0.024119095876812935
step: 140, loss: 0.0848955437541008
step: 150, loss: 0.026105405762791634
step: 160, loss: 0.0852922573685646
step: 170, loss: 0.0341547355055809
/usr/local/lib/python3.7/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=0.6993006993006994, f1=0.6790697674418605,
best_f1=0.7135922330097088
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01078171655535698
step: 10, loss: 0.09926334768533707
step: 20, loss: 0.007981089875102043
step: 30, loss: 0.0346548855304718
step: 40, loss: 0.005476702935993671
step: 50, loss: 0.05075956508517265
step: 60, loss: 0.0860941931605339
step: 70, loss: 0.009708035737276077
step: 80, loss: 0.1805901974439621
step: 90, loss: 0.10662203282117844
step: 100, loss: 0.12548862397670746
step: 110, loss: 0.07562781870365143
step: 120, loss: 0.048590365797281265
step: 130, loss: 0.0635543242096901
step: 140, loss: 0.0048276688903570175
step: 150, loss: 0.23831816017627716
step: 160, loss: 0.020463738590478897
step: 170, loss: 0.011381966061890125
/usr/local/lib/python3.7/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=0.6738544474393531, f1=0.70026525198939,
best_f1=0.7135922330097088
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

```
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13507190346717834
step: 10, loss: 0.00215840432792902
step: 20, loss: 0.04030141979455948
step: 30, loss: 0.2606649398803711
step: 40, loss: 0.008718535304069519
step: 50, loss: 0.0028141443617641926
step: 60, loss: 0.24139319360256195
step: 70, loss: 0.005369317252188921
step: 80, loss: 0.009897513315081596
step: 90, loss: 0.003791780211031437
step: 100, loss: 0.013137671165168285
step: 110, loss: 0.003135469974949956
step: 120, loss: 0.08432486653327942
step: 130, loss: 0.08432458341121674
step: 140, loss: 0.005789721850305796
step: 150, loss: 0.012553243897855282
step: 160, loss: 0.01445254497230053
step: 170, loss: 0.1506538689136505
/usr/local/lib/python3.7/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=0.7030878859857482, f1=0.6710816777041942,
best_f1=0.7135922330097088
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.028349293395876884
step: 10, loss: 0.0036438528914004564
step: 20, loss: 0.011244961060583591
step: 30, loss: 0.045726001262664795
step: 40, loss: 0.0022635932546108960
step: 50, loss: 0.02279164083302021
step: 60, loss: 0.0292336605489254
step: 70, loss: 0.005173825193196535
step: 80, loss: 0.010438934899866581
step: 90, loss: 0.04534563794732094
step: 100, loss: 0.0066126869060099125
step: 110, loss: 0.03647811710834503
step: 120, loss: 0.005496607162058353
step: 130, loss: 0.004590762313455343
step: 140, loss: 0.012667232193052769
step: 150, loss: 0.06976262480020523
step: 160, loss: 0.008359510451555252
step: 170, loss: 0.02176726423203945
/usr/local/lib/python3.7/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=0.6873508353221957, f1=0.6914153132250581,
best_f1=0.7135922330097088
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002186572179198265
step: 10, loss: 0.005979318171739578
step: 20, loss: 0.007830441929399967
step: 30, loss: 0.04662897065281868
step: 40, loss: 0.07027726620435715
step: 50, loss: 0.006503643933683634
step: 60, loss: 0.0050629605539143085
step: 70, loss: 0.1234280914068222
step: 80, loss: 0.01615181565284729
step: 90, loss: 0.0024676467292010784
step: 100, loss: 0.010249617509543896
step: 110, loss: 0.037427760660648346
step: 120, loss: 0.013643576763570309
step: 130, loss: 0.0006800565170124173
step: 140, loss: 0.0028958735056221485
step: 150, loss: 0.11422757059335709
step: 160, loss: 0.00943986140191555
step: 170, loss: 0.014903188683092594
/usr/local/lib/python3.7/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=0.6783042394014963, f1=0.6926829268292684,
best_f1=0.7135922330097088
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0017283431952819228
step: 10, loss: 0.0019483546493574977
step: 20, loss: 0.001757920952513814
step: 30, loss: 0.00042814016342163086
step: 40, loss: 0.001691766083240509
step: 50, loss: 0.0028404048644006252
step: 60, loss: 0.0007930818828754127
step: 70, loss: 0.09644650667905807
step: 80, loss: 0.002373490948230028
step: 90, loss: 0.03551719710230827
step: 100, loss: 0.0013830717653036118
step: 110, loss: 0.002063463907688856
step: 120, loss: 0.003945994656533003
step: 130, loss: 0.007879858836531639
step: 140, loss: 0.014978701248764992
step: 150, loss: 0.0028499187901616096
```

```
step: 160, loss: 0.0005892091430723667
step: 170, loss: 0.002732378663495183
/usr/local/lib/python3.7/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=0.6699029126213593, f1=0.6854460093896715,
best_f1=0.7135922330097088
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005722495261579752
step: 10, loss: 0.16957320272922516
step: 20, loss: 0.03840116411447525
step: 30, loss: 0.012776648625731468
step: 40, loss: 0.019223805516958237
step: 50, loss: 0.006290270481258631
step: 60, loss: 0.0019134259782731533
step: 70, loss: 0.007444364950060844
step: 80, loss: 0.0014965580776333809
step: 90, loss: 0.00046078822924755514
step: 100, loss: 0.0004698699340224266
step: 110, loss: 0.0010344600304961205
step: 120, loss: 0.21744151413440704
step: 130, loss: 0.0024906406179070473
step: 140, loss: 0.0073813581839203835
step: 150, loss: 0.0005764496163465083
step: 160, loss: 0.0005767575930804014
step: 170, loss: 0.001406301511451602
/usr/local/lib/python3.7/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=0.6733167082294264, f1=0.6921241050119331,
best_f1=0.7135922330097088
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006138970493339002
step: 10, loss: 0.0034161675721406937
step: 20, loss: 0.0020284440834075212
step: 30, loss: 0.0009283761028200388
step: 40, loss: 0.0005930141196586192
step: 50, loss: 0.0004564239934552461
step: 60, loss: 0.000434606074122712
step: 70, loss: 0.0889548510313034
step: 80, loss: 0.07870832830667496
step: 90, loss: 0.000439872412243858
step: 100, loss: 0.007066185120493174
step: 110, loss: 0.0010723511222749949
```

```
step: 120, loss: 0.0011838986538350582
step: 130, loss: 0.00435336260125041
step: 140, loss: 0.0009870856301859021
step: 150, loss: 0.0008648126968182623
step: 160, loss: 0.03263968974351883
step: 170, loss: 0.005404551979154348
/usr/local/lib/python3.7/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=0.6785714285714286, f1=0.7067307692307692,
best_f1=0.7135922330097088
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00045267364475876093
step: 10, loss: 0.0012000491842627525
step: 20, loss: 0.0016279099509119987
step: 30, loss: 0.0013702022843062878
step: 40, loss: 0.0003961588372476399
step: 50, loss: 0.0008367110276594758
step: 60, loss: 0.002320550847798586
step: 70, loss: 0.006494148168712854
step: 80, loss: 0.0006656351615674794
step: 90, loss: 0.00107954244595021
step: 100, loss: 0.00022233369236346334
step: 110, loss: 0.0007358766160905361
step: 120, loss: 0.0010015858570113778
step: 130, loss: 0.000831477576866746
step: 140, loss: 0.0010164843406528234
step: 150, loss: 0.0003658805217128247
step: 160, loss: 0.002920010592788458
step: 170, loss: 0.0006144240614958107
/usr/local/lib/python3.7/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=0.6761229314420805, f1=0.6909090909090909,
best_f1=0.7135922330097088
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00043147403630428016
step: 10, loss: 0.0005305611994117498
step: 20, loss: 0.000985272927209735
step: 30, loss: 0.0007386384531855583
step: 40, loss: 0.13315363228321075
step: 50, loss: 0.000905898807104677
step: 60, loss: 0.004285107366740704
step: 70, loss: 0.0005897236987948418
```

```
step: 80, loss: 0.0006450993241742253
step: 90, loss: 0.00028545389068312943
step: 100, loss: 0.0004042798245791346
step: 110, loss: 0.0011178087443113327
step: 120, loss: 0.0005121275316923857
step: 130, loss: 0.0005365686374716461
step: 140, loss: 0.0004726376791950315
step: 150, loss: 0.001865085680037737
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 160, loss: 0.03976484388113022
step: 170, loss: 0.0003629612911026925
/usr/local/lib/python3.7/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=0.6649616368286446, f1=0.698019801980198,
best_f1=0.7135922330097088
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02682551182806492
step: 10, loss: 0.0023801294155418873
step: 20, loss: 0.0045781852677464485
step: 30, loss: 0.028307868167757988
step: 40, loss: 0.0002527753240428865
step: 50, loss: 0.0003415956743992865
step: 60, loss: 0.0005003409460186958
step: 70, loss: 0.0006146466475911438
step: 80, loss: 0.0004298147396184504
step: 90, loss: 0.000909948255866766
step: 100, loss: 0.0019051338313147426
step: 110, loss: 0.00027872720966115594
step: 120, loss: 0.0011606727493926883
step: 130, loss: 0.03661429509520531
step: 140, loss: 0.0002954011724796146
step: 150, loss: 0.13186024129390717
step: 160, loss: 0.00025200832169502974
step: 170, loss: 0.00024375786597374827
/usr/local/lib/python3.7/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=0.6701030927835051, f1=0.6960784313725491,
best_f1=0.7135922330097088
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 [00:05, 374.30it/s]
load_f1 = 0.7482014388489209
real_f1 = 0.7400468384074942
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, 372.07it/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 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: 0.814489483833313
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.48781806230545044
step: 20, loss: 0.6145778894424438
step: 30, loss: 0.5015151500701904
step: 40, loss: 0.4429256021976471
step: 50, loss: 0.3443760871887207
step: 60, loss: 0.17493656277656555
step: 70, loss: 0.2202814519405365
step: 80, loss: 0.2898426353931427
step: 90, loss: 0.06689826399087906
step: 100, loss: 0.11768453568220139
step: 110, loss: 0.13516592979431152
step: 120, loss: 0.04825238138437271
step: 130, loss: 0.05126592889428139
step: 140, loss: 0.16812963783740997
step: 150, loss: 0.09894108772277832
step: 160, loss: 0.15776775777339935
step: 170, loss: 0.04915890097618103
step: 180, loss: 0.0133759630843997
step: 190, loss: 0.05878227949142456
step: 200, loss: 0.07085469365119934
step: 210, loss: 0.03249067813158035
step: 220, loss: 0.029665615409612656
step: 230, loss: 0.031024234369397163
/usr/local/lib/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=0.9345172031076582, f1=0.9382716049382717,
best_f1=0.9382716049382717
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09312225878238678
step: 10, loss: 0.04596473649144173
step: 20, loss: 0.07011017948389053
step: 30, loss: 0.009113159030675888

```
step: 40, loss: 0.05269140005111694
step: 50, loss: 0.11784442514181137
step: 60, loss: 0.10828869789838791
step: 70, loss: 0.18944597244262695
step: 80, loss: 0.019028257578611374
step: 90, loss: 0.00617460161447525
step: 100, loss: 0.20960800349712372
step: 110, loss: 0.09487877041101456
step: 120, loss: 0.03710990399122238
step: 130, loss: 0.21669110655784607
step: 140, loss: 0.13401702046394348
step: 150, loss: 0.057399407029151917
step: 160, loss: 0.03680230677127838
step: 170, loss: 0.023280341178178787
step: 180, loss: 0.010061162523925304
step: 190, loss: 0.09097343683242798
step: 200, loss: 0.05020042508840561
step: 210, loss: 0.12499090284109116
step: 220, loss: 0.007273358292877674
step: 230, loss: 0.0814865231513977
/usr/local/lib/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=0.9395973154362416, f1=0.9434389140271494,
best_f1=0.9434389140271494
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07387372106313705
step: 10, loss: 0.04363185912370682
step: 20, loss: 0.005947090219706297
step: 30, loss: 0.04849690571427345
step: 40, loss: 0.15038079023361206
step: 50, loss: 0.042236726731061935
step: 60, loss: 0.03578031063079834
step: 70, loss: 0.002798483707010746
step: 80, loss: 0.11811491847038269
step: 90, loss: 0.16171254217624664
step: 100, loss: 0.005251784343272448
step: 110, loss: 0.05005211383104324
step: 120, loss: 0.07108264416456223
step: 130, loss: 0.0038537951186299324
step: 140, loss: 0.001957807457074523
step: 150, loss: 0.0531463585793972
```

```
step: 160, loss: 0.05137402191758156
step: 170, loss: 0.0599118135869503
step: 180, loss: 0.0247525442391634
step: 190, loss: 0.007204726338386536
step: 200, loss: 0.00431451853364706
step: 210, loss: 0.027381982654333115
step: 220, loss: 0.005892922170460224
step: 230, loss: 0.10644371807575226
/usr/local/lib/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=0.953229398663697, f1=0.9430167597765362,
best_f1=0.9430167597765362
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02168833091855049
step: 10, loss: 0.01190017256885767
step: 20, loss: 0.003302921773865819
step: 30, loss: 0.0038575888611376286
step: 40, loss: 0.012498910538852215
step: 50, loss: 0.0032823635265231133
step: 60, loss: 0.0014145669993013144
step: 70, loss: 0.14791738986968994
step: 80, loss: 0.24825721979141235
step: 90, loss: 0.004677459131926298
step: 100, loss: 0.005616261623799801
step: 110, loss: 0.06428786367177963
step: 120, loss: 0.01292477734386921
step: 130, loss: 0.0021924260072410107
step: 140, loss: 0.008263214491307735
step: 150, loss: 0.001944119343534112
step: 160, loss: 0.0012762050610035658
step: 170, loss: 0.008128328248858452
step: 180, loss: 0.12417804449796677
step: 190, loss: 0.001468174858018756
step: 200, loss: 0.005914758425205946
step: 210, loss: 0.07733229547739029
step: 220, loss: 0.0026664386969059706
step: 230, loss: 0.07581069320440292
/usr/local/lib/python3.7/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=0.9508928571428572, f1=0.938132733408324,
```

```
best_f1=0.9430167597765362
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.018843134865164757
step: 10, loss: 0.014271107502281666
step: 20, loss: 0.0049428013153374195
step: 30, loss: 0.00028025006758980453
step: 40, loss: 0.007346214726567268
step: 50, loss: 0.000676060444675386
step: 60, loss: 0.0030631425324827433
step: 70, loss: 0.0004825841751880944
step: 80, loss: 0.0014323705108836293
step: 90, loss: 0.001632949337363243
step: 100, loss: 0.008886642754077911
step: 110, loss: 0.0021135490387678146
step: 120, loss: 0.032504212111234665
step: 130, loss: 0.048409052193164825
step: 140, loss: 0.002185484394431114
step: 150, loss: 0.006427498999983072
step: 160, loss: 0.15850242972373962
step: 170, loss: 0.017152005806565285
step: 180, loss: 0.004396806936711073
step: 190, loss: 0.0017112703062593937
step: 200, loss: 0.0030371826142072678
step: 210, loss: 0.0023610065691173077
step: 220, loss: 0.0034641982056200504
step: 230, loss: 0.0013139168731868267
/usr/local/lib/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=0.9552572706935123, f1=0.9461883408071748,
best_f1=0.9461883408071748
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0036174370907247066
step: 10, loss: 0.06631707400083542
step: 20, loss: 0.011328352615237236
step: 30, loss: 0.004946187138557434
step: 40, loss: 0.017170144245028496
step: 50, loss: 0.0038986122235655785
step: 60, loss: 0.02239209972321987
step: 70, loss: 0.00866083987057209
```

```
step: 80, loss: 0.0033191777765750885
step: 90, loss: 0.00149985880125314
step: 100, loss: 0.003917364869266748
step: 110, loss: 0.09822045266628265
step: 120, loss: 0.0011497812811285257
step: 130, loss: 0.0011015981435775757
step: 140, loss: 0.0017664044862613082
step: 150, loss: 0.0006781133124604821
step: 160, loss: 0.00171833171043545
step: 170, loss: 0.0010141978273168206
step: 180, loss: 0.008912323042750359
step: 190, loss: 0.0113186314702034
step: 200, loss: 0.0037631921004503965
step: 210, loss: 0.0049108886159956455
step: 220, loss: 0.0007829729001969099
step: 230, loss: 0.05152969807386398
/usr/local/lib/python3.7/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=0.9467849223946784, f1=0.936026936026936,
best_f1=0.9461883408071748
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1842779815196991
step: 10, loss: 0.0019002215703949332
step: 20, loss: 0.0018070163205265999
step: 30, loss: 0.0162779800593853
step: 40, loss: 0.0022824800107628107
step: 50, loss: 0.0007632087799720466
step: 60, loss: 0.03466685116291046
step: 70, loss: 0.063769318163394933
step: 80, loss: 0.0026289261877536774
step: 90, loss: 0.005540068726986647
step: 100, loss: 0.0009424446034245193
step: 110, loss: 0.0015039379941299558
step: 120, loss: 0.012878300622105598
step: 130, loss: 0.0021747665014117956
step: 140, loss: 0.0008224168559536338
step: 150, loss: 0.0016475036973133683
step: 160, loss: 0.00039900446427054703
step: 170, loss: 0.0006072286632843316
step: 180, loss: 0.000999621581286192
step: 190, loss: 0.0006927552167326212
step: 200, loss: 0.005892452318221331
step: 210, loss: 0.0005156842526048422
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 220, loss: 0.03137669712305069
```

```
step: 230, loss: 0.04280814900994301
/usr/local/lib/python3.7/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=0.9364035087719298, f1=0.9377777777777777,
best_f1=0.9461883408071748
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03427577391266823
step: 10, loss: 0.0593993552029913284
step: 20, loss: 0.0022849508095532656
step: 30, loss: 0.0017271110555157065
step: 40, loss: 0.0014912192709743977
step: 50, loss: 0.017460854724049568
step: 60, loss: 0.0007821201579645276
step: 70, loss: 0.0042071957141160965
step: 80, loss: 0.009788758279234171
step: 90, loss: 0.003249618923291564
step: 100, loss: 0.0010350922821089625
step: 110, loss: 0.010656801983714104
step: 120, loss: 0.00019407464424148202
step: 130, loss: 0.023637397214770317
step: 140, loss: 0.00028941419441252947
step: 150, loss: 0.0003937128640245646
step: 160, loss: 0.0008830523584038019
step: 170, loss: 0.00031563377706333995
step: 180, loss: 0.004249304998666048
step: 190, loss: 0.006661190651357174
step: 200, loss: 0.01505704503506422
step: 210, loss: 0.009896811097860336
step: 220, loss: 0.0017104515573009849
step: 230, loss: 0.19974835216999054
/usr/local/lib/python3.7/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=0.9480812641083521, f1=0.9350057012542761,
best_f1=0.9461883408071748
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00018136498692911118
step: 10, loss: 0.0004314029647503048
step: 20, loss: 0.004871074575930834
step: 30, loss: 0.0007293711532838643
step: 40, loss: 0.00023629871429875493
step: 50, loss: 0.0001951568410731852
step: 60, loss: 0.0004413228889461607
```

```
step: 70, loss: 0.0006013137171976268
step: 80, loss: 0.0209847129881382
step: 90, loss: 0.0006841900758445263
step: 100, loss: 0.004429434891790152
step: 110, loss: 0.0015151159267000854
step: 120, loss: 0.11540920287370682
step: 130, loss: 0.0007137209759093821
step: 140, loss: 0.08536972850561142
step: 150, loss: 0.0001819443132262677
step: 160, loss: 0.0012449441710487008
step: 170, loss: 0.0008118193945847452
step: 180, loss: 0.00045070311171002686
step: 190, loss: 0.0004027529212180525
step: 200, loss: 0.00033341243397444487
step: 210, loss: 0.0004629569302778691
step: 220, loss: 0.00028603570535779
step: 230, loss: 0.0012782547855749726
/usr/local/lib/python3.7/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=0.9443813847900113, f1=0.9260969976905312,
best_f1=0.9461883408071748
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00038023124216124415
step: 10, loss: 0.00029298049048520625
step: 20, loss: 0.0005608895444311202
step: 30, loss: 0.00036277694744057953
step: 40, loss: 0.00010263931471854448
step: 50, loss: 0.0018161280313506722
step: 60, loss: 0.05529871582984924
step: 70, loss: 0.004636180587112904
step: 80, loss: 0.00413444684818387
step: 90, loss: 0.00023643262102268636
step: 100, loss: 0.00019456727022770792
step: 110, loss: 0.00017566238238941878
step: 120, loss: 0.0035341728944331408
step: 130, loss: 0.006881463807076216
step: 140, loss: 0.024253062903881073
step: 150, loss: 0.0002908961905632168
step: 160, loss: 0.001187468646094203
step: 170, loss: 0.00043741019908338785
step: 180, loss: 0.0013614136260002851
step: 190, loss: 0.000800392939709127
step: 200, loss: 0.000234966617426835
step: 210, loss: 0.0005636237328872085
step: 220, loss: 0.0006354405195452273
```

```
step: 230, loss: 0.0013088478008285165
/usr/local/lib/python3.7/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=0.947250280583614, f1=0.9389140271493213,
best_f1=0.9461883408071748
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0004198603564873338
step: 10, loss: 0.004118913318961859
step: 20, loss: 0.0002315440506208688
step: 30, loss: 0.0006447188789024949
step: 40, loss: 0.0027423801366239786
step: 50, loss: 0.001040876260958612
step: 60, loss: 0.0003576611343305558
step: 70, loss: 0.004111029207706451
step: 80, loss: 0.00034814196988008916
step: 90, loss: 0.0927751436829567
step: 100, loss: 0.0021873670630156994
step: 110, loss: 0.014122658409178257
step: 120, loss: 0.0011070016771554947
step: 130, loss: 0.0004433701396919787
step: 140, loss: 0.0001281259028473869
step: 150, loss: 0.0001215523443534039
step: 160, loss: 0.00017043172556441277
step: 170, loss: 0.0008681650506332517
step: 180, loss: 0.00032899691723287106
step: 190, loss: 0.0002501935523469001
step: 200, loss: 0.00017400133947376162
step: 210, loss: 0.0001272126828553155
step: 220, loss: 0.0005011051543988287
step: 230, loss: 0.0009086175705306232
/usr/local/lib/python3.7/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=0.9440000000000001, f1=0.916083916083916,
best_f1=0.9461883408071748
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003549036046024412
step: 10, loss: 0.0018283799290657043
step: 20, loss: 0.00018217445176560432
step: 30, loss: 0.0002489920298103243
step: 40, loss: 0.00010961965745082125
step: 50, loss: 0.00019329015049152076
step: 60, loss: 0.014678620733320713
```

```
step: 70, loss: 0.00030910249915905297
step: 80, loss: 9.587375825503841e-05
step: 90, loss: 0.00014131487114354968
step: 100, loss: 0.00031166159897111356
step: 110, loss: 0.00029824141529388726
step: 120, loss: 0.0003055720590054989
step: 130, loss: 0.00017644606123212725
step: 140, loss: 0.00015948963118717074
step: 150, loss: 7.172362529672682e-05
step: 160, loss: 9.150299592874944e-05
step: 170, loss: 0.00017430954903829843
step: 180, loss: 0.00012627728574443609
step: 190, loss: 0.00012244636309333146
step: 200, loss: 0.0005643479526042938
step: 210, loss: 0.00027329556178301573
step: 220, loss: 0.03390584513545036
step: 230, loss: 0.00026270089438185096
/usr/local/lib/python3.7/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=0.9426048565121413, f1=0.9473684210526316,
best_f1=0.9461883408071748
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.217434853781015e-05
step: 10, loss: 9.328242595074698e-05
step: 20, loss: 0.0001902538351714611
step: 30, loss: 0.024083368480205536
step: 40, loss: 0.0003345845325384289
step: 50, loss: 0.00014077601372264326
step: 60, loss: 0.00010255259985569865
step: 70, loss: 0.00017633329844102263
step: 80, loss: 0.004467830527573824
step: 90, loss: 6.996157026151195e-05
step: 100, loss: 0.00018755388737190515
step: 110, loss: 0.00010629291500663385
step: 120, loss: 0.00021175383881200105
step: 130, loss: 0.00023204507306218147
step: 140, loss: 0.002609068527817726
step: 150, loss: 0.00817249808460474
step: 160, loss: 0.00017737555026542395
step: 170, loss: 0.0001362945040455088
step: 180, loss: 0.00022174607147462666
step: 190, loss: 0.0001198849204229191
step: 200, loss: 0.00012177741882624105
step: 210, loss: 0.00011519790132297203
step: 220, loss: 8.043617708608508e-05
```

```
step: 230, loss: 8.586268813814968e-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=0.9423728813559321, f1=0.9331797235023043,
best_f1=0.9461883408071748
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00010121092782355845
step: 10, loss: 7.805632776580751e-05
step: 20, loss: 0.000440483505371958
step: 30, loss: 0.00043172293226234615
step: 40, loss: 8.26369141577743e-05
step: 50, loss: 0.000250777491601184
step: 60, loss: 0.00021370782633312047
step: 70, loss: 9.947707440005615e-05
step: 80, loss: 0.00016048431280069053
step: 90, loss: 8.929677278501913e-05
step: 100, loss: 0.0004242270370014012
step: 110, loss: 0.003277908079326153
step: 120, loss: 0.013033553026616573
step: 130, loss: 0.00015380473632831126
step: 140, loss: 8.577782864449546e-05
step: 150, loss: 0.00011335036106174812
step: 160, loss: 6.962811312405393e-05
step: 170, loss: 0.00012218556366860867
step: 180, loss: 0.00026012645685113966
step: 190, loss: 0.0007045022211968899
step: 200, loss: 6.601051427423954e-05
step: 210, loss: 0.0013154015177860856
step: 220, loss: 0.00018242215446662158
step: 230, loss: 4.8850375605979934e-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=0.9425028184892897, f1=0.9325714285714286,
best_f1=0.9461883408071748
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00012553489068523305
step: 10, loss: 0.000165329227456823
step: 20, loss: 0.0001674996456131339
step: 30, loss: 0.0010954377939924598
step: 40, loss: 0.00034872785909101367
step: 50, loss: 0.0001577125076437369
step: 60, loss: 4.591152537614107e-05
```

```
step: 70, loss: 0.00011567794717848301
step: 80, loss: 0.00015750553575344384
step: 90, loss: 0.00015372515190392733
step: 100, loss: 0.00015356627409346402
step: 110, loss: 0.00011476752115413547
step: 120, loss: 0.00024111766833811998
step: 130, loss: 0.0009544765343889594
step: 140, loss: 0.005434723105281591
step: 150, loss: 0.00013531824515666813
step: 160, loss: 0.0006853314116597176
step: 170, loss: 6.568078242707998e-05
step: 180, loss: 0.00014167727204039693
step: 190, loss: 0.0037640095688402653
step: 200, loss: 8.812181476969272e-05
step: 210, loss: 0.0023857520427554846
step: 220, loss: 0.027688302099704742
step: 230, loss: 0.0004663541913032532
/usr/local/lib/python3.7/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=0.9427609427609427, f1=0.937428896473265,
best_f1=0.9461883408071748
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 [00:08, 290.12it/s]
load_f1 = 0.9541899441340782
real_f1 = 0.9541899441340782
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, 360.13it/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 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: 0.7996423840522766
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4531988799571991
step: 20, loss: 0.5181674361228943
step: 30, loss: 0.47831034660339355
step: 40, loss: 0.4772384762763977
step: 50, loss: 0.43809208273887634
step: 60, loss: 0.480212926864624
step: 70, loss: 0.19481585919857025
step: 80, loss: 0.37870287895202637
step: 90, loss: 0.39151132106781006
step: 100, loss: 0.24875375628471375
step: 110, loss: 0.179433211684227
step: 120, loss: 0.12311061471700668
step: 130, loss: 0.03741825371980667
step: 140, loss: 0.29342958331108093
```

```
step: 150, loss: 0.08082432299852371
step: 160, loss: 0.16619153320789337
step: 170, loss: 0.34405767917633057
step: 180, loss: 0.18068313598632812
step: 190, loss: 0.08044296503067017
step: 200, loss: 0.15360690653324127
step: 210, loss: 0.11788473278284073
step: 220, loss: 0.07096456736326218
step: 230, loss: 0.08059045672416687
step: 240, loss: 0.09163957834243774
step: 250, loss: 0.10735619068145752
step: 260, loss: 0.03097505308687687
step: 270, loss: 0.022827940061688423
step: 280, loss: 0.2633565664291382
step: 290, loss: 0.11020497232675552
step: 300, loss: 0.1555689573287964
step: 310, loss: 0.06299659609794617
step: 320, loss: 0.12660396099090576
step: 330, loss: 0.1233941838145256
step: 340, loss: 0.2544216513633728
step: 350, loss: 0.05846628174185753
step: 360, loss: 0.09937279671430588
step: 370, loss: 0.1422802358865738
step: 380, loss: 0.32564279437065125
step: 390, loss: 0.13509249687194824
step: 400, loss: 0.03043108619749546
step: 410, loss: 0.11129089444875717
step: 420, loss: 0.0990593358874321
step: 430, loss: 0.20507925748825073
step: 440, loss: 0.13442477583885193
step: 450, loss: 0.07429838925600052
step: 460, loss: 0.02586640976369381
step: 470, loss: 0.1376579999923706
step: 480, loss: 0.23634764552116394
step: 490, loss: 0.06236817315220833
step: 500, loss: 0.033296599984169006
step: 510, loss: 0.145770862698555
step: 520, loss: 0.12511411309242249
step: 530, loss: 0.037904445081949234
/usr/local/lib/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=0.9021937842778793, f1=0.8985773290500229,
best_f1=0.8985773290500229
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09967906028032303
step: 10, loss: 0.22516654431819916
step: 20, loss: 0.09377527981996536
step: 30, loss: 0.07763013243675232
step: 40, loss: 0.011509106494486332
step: 50, loss: 0.08416309207677841
step: 60, loss: 0.13866916298866272
step: 70, loss: 0.16718167066574097
step: 80, loss: 0.034490592777729034
step: 90, loss: 0.12437502294778824
step: 100, loss: 0.29771339893341064
step: 110, loss: 0.11125250160694122
step: 120, loss: 0.06341274827718735
step: 130, loss: 0.08511684834957123
step: 140, loss: 0.06989709287881851
step: 150, loss: 0.06103016436100006
step: 160, loss: 0.0423630066215992
step: 170, loss: 0.21955378353595734
step: 180, loss: 0.08263280987739563
step: 190, loss: 0.014768457971513271
step: 200, loss: 0.06735427677631378
step: 210, loss: 0.027165982872247696
step: 220, loss: 0.14469054341316223
step: 230, loss: 0.12923437356948853
step: 240, loss: 0.12241664528846741
step: 250, loss: 0.0850035771727562
step: 260, loss: 0.09146285057067871
step: 270, loss: 0.19106930494308472
step: 280, loss: 0.23906055092811584
step: 290, loss: 0.07488849759101868
step: 300, loss: 0.07577686011791229
step: 310, loss: 0.11461343616247177
step: 320, loss: 0.16370168328285217
step: 330, loss: 0.06720801442861557
step: 340, loss: 0.008895430713891983
step: 350, loss: 0.17046263813972473
step: 360, loss: 0.0758349671959877
step: 370, loss: 0.0589132159948349
step: 380, loss: 0.14950847625732422
step: 390, loss: 0.037243977189064026
step: 400, loss: 0.12677551805973053
step: 410, loss: 0.00222926028072834
step: 420, loss: 0.12465258687734604
step: 430, loss: 0.03930887579917908
step: 440, loss: 0.017550883814692497
```

```
step: 450, loss: 0.1103738471865654
step: 460, loss: 0.15866313874721527
step: 470, loss: 0.024800147861242294
step: 480, loss: 0.38800978660583496
step: 490, loss: 0.040572915226221085
step: 500, loss: 0.01645093783736229
step: 510, loss: 0.06573927402496338
step: 520, loss: 0.044113464653491974
step: 530, loss: 0.06922230124473572
/usr/local/lib/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=0.9154275092936803, f1=0.9039704524469068,
best_f1=0.9039704524469068
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04170197993516922
step: 10, loss: 0.05267787724733353
step: 20, loss: 0.299456924200058
step: 30, loss: 0.1531735360622406
step: 40, loss: 0.057415809482336044
step: 50, loss: 0.06906771659851074
step: 60, loss: 0.0064699179492890835
step: 70, loss: 0.031685199588537216
step: 80, loss: 0.06531237065792084
step: 90, loss: 0.3142153322696686
step: 100, loss: 0.09282438457012177
step: 110, loss: 0.01836032047867775
step: 120, loss: 0.015150632709264755
step: 130, loss: 0.02909386157989502
step: 140, loss: 0.08229197561740875
step: 150, loss: 0.005517967510968447
step: 160, loss: 0.04853542521595955
step: 170, loss: 0.007920455187559128
step: 180, loss: 0.11575141549110413
step: 190, loss: 0.012545754201710224
step: 200, loss: 0.03304686397314072
step: 210, loss: 0.1720399409532547
step: 220, loss: 0.027444183826446533
step: 230, loss: 0.011568532325327396
step: 240, loss: 0.08452831953763962
step: 250, loss: 0.03154844045639038
step: 260, loss: 0.0029951445758342743
```

```
step: 270, loss: 0.0377519465982914
step: 280, loss: 0.02571670524775982
step: 290, loss: 0.08456654101610184
step: 300, loss: 0.0683378130197525
step: 310, loss: 0.0967329815030098
step: 320, loss: 0.1041552945971489
step: 330, loss: 0.08043544739484787
step: 340, loss: 0.006338839884847403
step: 350, loss: 0.13880249857902527
step: 360, loss: 0.012198081240057945
step: 370, loss: 0.020007692277431488
step: 380, loss: 0.006063760723918676
step: 390, loss: 0.047482337802648544
step: 400, loss: 0.028260059554712677
step: 410, loss: 0.01569652184844017
step: 420, loss: 0.08313288539648056
step: 430, loss: 0.06184730306267738
step: 440, loss: 0.042618878185749054
step: 450, loss: 0.13794319331645966
step: 460, loss: 0.21523648500442505
step: 470, loss: 0.011034335941076279
step: 480, loss: 0.015758804976940155
step: 490, loss: 0.0063642049208283424
step: 500, loss: 0.08859975636005402
step: 510, loss: 0.080877386033535
step: 520, loss: 0.0035205066669732332
step: 530, loss: 0.02686295285820961
/usr/local/lib/python3.7/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=0.9049394221808016, f1=0.8969411764705882,
best_f1=0.9039704524469068
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.022679559886455536
step: 10, loss: 0.10009033977985382
step: 20, loss: 0.048448387533426285
step: 30, loss: 0.049630388617515564
step: 40, loss: 0.08614019304513931
step: 50, loss: 0.005322567652910948
step: 60, loss: 0.022777339443564415
step: 70, loss: 0.00982017070055008
step: 80, loss: 0.03871545195579529
step: 90, loss: 0.08268319070339203
step: 100, loss: 0.0029233302921056747
step: 110, loss: 0.005875647068023682
step: 120, loss: 0.05305924266576767
```

```
step: 130, loss: 0.06675305962562561
step: 140, loss: 0.018150314688682556
step: 150, loss: 0.0018223256338387728
step: 160, loss: 0.012516687624156475
step: 170, loss: 0.002706179628148675
step: 180, loss: 0.011110854335129261
step: 190, loss: 0.008068357594311237
step: 200, loss: 0.0016926862299442291
step: 210, loss: 0.022151025012135506
step: 220, loss: 0.009807681664824486
step: 230, loss: 0.2869276702404022
step: 240, loss: 0.007965579628944397
step: 250, loss: 0.08467801660299301
step: 260, loss: 0.01657911203801632
step: 270, loss: 0.043173860758543015
step: 280, loss: 0.024567391723394394
step: 290, loss: 0.12247809022665024
step: 300, loss: 0.013722556643188
step: 310, loss: 0.010951833799481392
step: 320, loss: 0.09152722358703613
step: 330, loss: 0.06959055364131927
step: 340, loss: 0.009050016291439533
step: 350, loss: 0.016076255589723587
step: 360, loss: 0.012883871793746948
step: 370, loss: 0.004603770095854998
step: 380, loss: 0.0036002707201987505
step: 390, loss: 0.004201050382107496
step: 400, loss: 0.028482254594564438
step: 410, loss: 0.015498366206884384
step: 420, loss: 0.05132339149713516
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 430, loss: 0.028038162738084793
step: 440, loss: 0.11277024447917938
step: 450, loss: 0.006867324002087116
step: 460, loss: 0.008198808878660202
step: 470, loss: 0.004507246892899275
step: 480, loss: 0.05769260227680206
step: 490, loss: 0.04425814747810364
step: 500, loss: 0.018195437267422676
step: 510, loss: 0.02143595553934574
step: 520, loss: 0.12350798398256302
step: 530, loss: 0.004333063960075378
/usr/local/lib/python3.7/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=0.9046500704556129, f1=0.8944392082940621,
best_f1=0.9039704524469068
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

```
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003550855442881584
step: 10, loss: 0.05871516093611717
step: 20, loss: 0.057234372943639755
step: 30, loss: 0.008429497480392456
step: 40, loss: 0.049983371049165726
step: 50, loss: 0.021753134205937386
step: 60, loss: 0.010459651239216328
step: 70, loss: 0.001984945498406887
step: 80, loss: 0.001101384754292667
step: 90, loss: 0.009941254742443562
step: 100, loss: 0.0024110875092446804
step: 110, loss: 0.001979703549295664
step: 120, loss: 0.0383453443646431
step: 130, loss: 0.013727056793868542
step: 140, loss: 0.01629684306681156
step: 150, loss: 0.03771762177348137
step: 160, loss: 0.16033999621868134
step: 170, loss: 0.009970476850867271
step: 180, loss: 0.08020877093076706
step: 190, loss: 0.015425290912389755
step: 200, loss: 0.0038672632072120905
step: 210, loss: 0.0030196458101272583
step: 220, loss: 0.0023867327254265547
step: 230, loss: 0.010620261542499065
step: 240, loss: 0.05411624163389206
step: 250, loss: 0.002079189755022526
step: 260, loss: 0.020764803513884544
step: 270, loss: 0.007960702292621136
step: 280, loss: 0.06865067780017853
step: 290, loss: 0.0006597767933271825
step: 300, loss: 0.015359065495431423
step: 310, loss: 0.0007404147181659937
step: 320, loss: 0.00537307932972908
step: 330, loss: 0.11750238388776779
step: 340, loss: 0.00276299100369215
step: 350, loss: 0.001981459790840745
step: 360, loss: 0.04987072944641113
step: 370, loss: 0.005158618092536926
step: 380, loss: 0.025466330349445343
step: 390, loss: 0.015002626925706863
step: 400, loss: 0.004623066168278456
step: 410, loss: 0.002571123419329524
step: 420, loss: 0.010251132771372795
step: 430, loss: 0.019240915775299072
step: 440, loss: 0.046277277171611786
step: 450, loss: 0.014718924649059772
```

```
step: 460, loss: 0.0015772258630022407
step: 470, loss: 0.009952700696885586
step: 480, loss: 0.014332267455756664
step: 490, loss: 0.008072856813669205
step: 500, loss: 0.006102705840021372
step: 510, loss: 0.007605484686791897
step: 520, loss: 0.0034924191422760487
step: 530, loss: 0.006694120820611715
/usr/local/lib/python3.7/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=0.9125939849624061, f1=0.8983855650522317,
best_f1=0.9039704524469068
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.024289576336741447
step: 10, loss: 0.00696193240582943
step: 20, loss: 0.0017232733080163598
step: 30, loss: 0.04646644368767738
step: 40, loss: 0.052797671407461166
step: 50, loss: 0.0013156816130504012
step: 60, loss: 0.003655944252386689
step: 70, loss: 0.0310970488935709
step: 80, loss: 0.0003664682153612375
step: 90, loss: 0.01587996818125248
step: 100, loss: 0.10398095101118088
step: 110, loss: 0.00576772540807724
step: 120, loss: 0.00966945756226778
step: 130, loss: 0.004439039155840874
step: 140, loss: 0.02828352525830269
step: 150, loss: 0.008091249503195286
step: 160, loss: 0.04032314568758011
step: 170, loss: 0.0007627442246302962
step: 180, loss: 0.00037191619048826396
step: 190, loss: 0.004821273032575846
step: 200, loss: 0.0019694154616445303
step: 210, loss: 0.0009998271707445383
step: 220, loss: 0.0008615648839622736
step: 230, loss: 0.000666275096591562
step: 240, loss: 0.0251960139721632
step: 250, loss: 0.002705363556742668
step: 260, loss: 0.0005409325240179896
step: 270, loss: 0.00043234415352344513
step: 280, loss: 0.001438722014427185
step: 290, loss: 0.001184886903502047
step: 300, loss: 0.00045690464321523905
step: 310, loss: 0.018925292417407036
```

```
step: 320, loss: 0.00017548275354783982
step: 330, loss: 0.0005540476995520294
step: 340, loss: 0.07381068170070648
step: 350, loss: 0.00432970467954874
step: 360, loss: 0.003494270844385028
step: 370, loss: 0.0018351242179051042
step: 380, loss: 0.001133721205405891
step: 390, loss: 0.008134541101753712
step: 400, loss: 0.0001907299447339028
step: 410, loss: 0.0004427308449521661
step: 420, loss: 0.007844441570341587
step: 430, loss: 0.0025843337643891573
step: 440, loss: 0.0195468682795763
step: 450, loss: 0.001417031278833747
step: 460, loss: 0.20401601493358612
step: 470, loss: 0.24887314438819885
step: 480, loss: 0.0039696418680250645
step: 490, loss: 0.0025980875361710787
step: 500, loss: 0.004809664096683264
step: 510, loss: 0.0029837065376341343
step: 520, loss: 0.00852635595947504
step: 530, loss: 0.03386378660798073
/usr/local/lib/python3.7/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=0.9063670411985018, f1=0.8990999526290857,
best_f1=0.9039704524469068
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08862873166799545
step: 10, loss: 0.008252574130892754
step: 20, loss: 0.049535609781742096
step: 30, loss: 0.0064278095960617065
step: 40, loss: 0.002348031383007765
step: 50, loss: 0.003014717483893037
step: 60, loss: 0.002811809303238988
step: 70, loss: 0.0012372517958283424
step: 80, loss: 0.013505881652235985
step: 90, loss: 0.00022432452533394098
step: 100, loss: 0.0005119108245708048
step: 110, loss: 0.0003486491914372891
step: 120, loss: 0.0012875903630629182
step: 130, loss: 0.0005173508543521166
step: 140, loss: 0.000590306066442281
step: 150, loss: 0.0008002883987501264
step: 160, loss: 0.0005533329676836729
step: 170, loss: 0.01966038905084133
```

```
step: 180, loss: 0.0002035719808191061
step: 190, loss: 0.001960823079571128
step: 200, loss: 0.03135065361857414
step: 210, loss: 0.034702736884355545
step: 220, loss: 0.008197850547730923
step: 230, loss: 0.001016082358546555
step: 240, loss: 0.0019936009775847197
step: 250, loss: 0.10510482639074326
step: 260, loss: 0.0010555546032264829
step: 270, loss: 0.00025598282809369266
step: 280, loss: 0.01253444328904152
step: 290, loss: 0.012491128407418728
step: 300, loss: 0.0011837314814329147
step: 310, loss: 0.00023096306540537626
step: 320, loss: 0.010133779607713223
step: 330, loss: 0.003028753213584423
step: 340, loss: 0.008813809603452682
step: 350, loss: 0.001046965247951448
step: 360, loss: 0.002582848072052002
step: 370, loss: 0.045369476079940796
step: 380, loss: 0.019791567769812107
step: 390, loss: 0.0107661047950387
step: 400, loss: 0.0023952166084200144
step: 410, loss: 0.002234665909782052
step: 420, loss: 0.004018580541014671
step: 430, loss: 0.00015614127914886922
step: 440, loss: 0.00027398421661928296
step: 450, loss: 0.00136516522616148
step: 460, loss: 0.0014940116088837385
step: 470, loss: 0.0524381548166275
step: 480, loss: 0.00224848510697484
step: 490, loss: 0.0009244962129741907
step: 500, loss: 0.00015742868708912283
step: 510, loss: 0.0017735975561663508
step: 520, loss: 0.3349100947380066
step: 530, loss: 0.0004130580055061728
/usr/local/lib/python3.7/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=0.9085768143261074, f1=0.9009009009009009,
best_f1=0.9039704524469068
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005889153690077364
step: 10, loss: 0.011127650737762451
step: 20, loss: 0.004669674672186375
step: 30, loss: 0.05397910624742508
```

```
step: 40, loss: 0.009778554551303387
step: 50, loss: 0.001115144114010036
step: 60, loss: 0.0014765691012144089
step: 70, loss: 0.0009861803846433759
step: 80, loss: 0.00022310821805149317
step: 90, loss: 0.004011086653918028
step: 100, loss: 0.00036065600579604506
step: 110, loss: 0.0011038730153813958
step: 120, loss: 0.0019381206948310137
step: 130, loss: 0.0003371835045982152
step: 140, loss: 8.934747893363237e-05
step: 150, loss: 0.0006329749012365937
step: 160, loss: 0.0013044241350144148
step: 170, loss: 0.003687160322442651
step: 180, loss: 0.001660062000155449
step: 190, loss: 0.005554789677262306
step: 200, loss: 0.0019148490391671658
step: 210, loss: 0.08468234539031982
step: 220, loss: 0.000914761156309396
step: 230, loss: 0.000696048024110496
step: 240, loss: 0.0004881521745119244
step: 250, loss: 0.0019634768832474947
step: 260, loss: 0.0026773724239319563
step: 270, loss: 0.0003385945165064186
step: 280, loss: 0.01328683365136385
step: 290, loss: 0.314879834651947
step: 300, loss: 0.00012992304982617497
step: 310, loss: 0.08548165112733841
step: 320, loss: 0.0002222400944447145
step: 330, loss: 0.0030393691267818213
step: 340, loss: 0.003966981079429388
step: 350, loss: 0.014482099562883377
step: 360, loss: 0.0002835097548086196
step: 370, loss: 0.00163958128541708
step: 380, loss: 0.0037162508815526962
step: 390, loss: 0.0011029393645003438
step: 400, loss: 0.03247763216495514
step: 410, loss: 0.0006472698296420276
step: 420, loss: 0.000313727417960763
step: 430, loss: 0.00016431727271992713
step: 440, loss: 0.0011086463928222656
step: 450, loss: 0.00032464347896166146
step: 460, loss: 0.013047423213720322
step: 470, loss: 0.0005118386470712721
step: 480, loss: 0.0013788407668471336
step: 490, loss: 0.000580630439799279
step: 500, loss: 0.000280822190688923
step: 510, loss: 0.00019966778927482665
```

```
step: 520, loss: 0.0003803520812653005
step: 530, loss: 0.0001615330984350294
/usr/local/lib/python3.7/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=0.9017773620205801, f1=0.898876404494382,
best_f1=0.9039704524469068
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00023553622304461896
step: 10, loss: 0.00027241290081299239
step: 20, loss: 0.000416945549659431
step: 30, loss: 0.08493176102638245
step: 40, loss: 0.11251581460237503
step: 50, loss: 0.00029316041036050898
step: 60, loss: 0.005573255941271782
step: 70, loss: 0.0015247628325596452
step: 80, loss: 0.00015883291780482978
step: 90, loss: 0.01369550358504057
step: 100, loss: 0.00147941242903471
step: 110, loss: 0.00013731281796935946
step: 120, loss: 0.00016761691949795932
step: 130, loss: 0.004706798121333122
step: 140, loss: 0.004767163656651974
step: 150, loss: 8.174929826054722e-05
step: 160, loss: 0.00044643902219831944
step: 170, loss: 0.0007491235737688839
step: 180, loss: 9.779255196917802e-05
step: 190, loss: 0.0013708164915442467
step: 200, loss: 0.00023802609939593822
step: 210, loss: 0.00012008116027573124
step: 220, loss: 0.002963021397590637
step: 230, loss: 0.0005066500161774457
step: 240, loss: 0.0007420388865284622
step: 250, loss: 0.0008228387450799346
step: 260, loss: 0.0013245587470009923
step: 270, loss: 0.010014832019805908
step: 280, loss: 0.0002398222277406603
step: 290, loss: 0.00010879441106226295
step: 300, loss: 0.0019286139868199825
step: 310, loss: 0.0003485619090497494
step: 320, loss: 0.0006553790881298482
step: 330, loss: 0.0003354674845468253
step: 340, loss: 0.00011880089004989713
step: 350, loss: 0.00012716690253000706
step: 360, loss: 0.00459745479747653
step: 370, loss: 0.012061882764101028
```

```
step: 380, loss: 0.0004044279339723289
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 390, loss: 0.00018631271086633205
step: 400, loss: 0.003473997814580798
step: 410, loss: 0.0010324701434001327
step: 420, loss: 0.0009461604058742523
step: 430, loss: 0.00012570508988574147
step: 440, loss: 9.821002458920702e-05
step: 450, loss: 9.979523747460917e-05
step: 460, loss: 0.021887360140681267
step: 470, loss: 0.00012560366303659976
step: 480, loss: 0.001393738086335361
step: 490, loss: 0.0006316562066785991
step: 500, loss: 0.0006401462596841156
step: 510, loss: 0.0002446482831146568
step: 520, loss: 0.0001377338485326618
step: 530, loss: 0.0005770967109128833
/usr/local/lib/python3.7/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=0.9087505849321479, f1=0.9049858889934148,
best_f1=0.9039704524469068
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0039022478740662336
step: 10, loss: 9.433190280105919e-05
step: 20, loss: 0.0013365333434194326
step: 30, loss: 0.00015147047815844417
step: 40, loss: 0.0038029206916689873
step: 50, loss: 9.884469181997702e-05
step: 60, loss: 0.00027830631006509066
step: 70, loss: 0.0004419922479428351
step: 80, loss: 0.00029319257009774446
step: 90, loss: 0.00020277457952033728
step: 100, loss: 0.0017030220478773117
step: 110, loss: 0.0032561644911766052
step: 120, loss: 0.0004919233033433557
step: 130, loss: 0.021349826827645302
step: 140, loss: 0.00014087988529354334
step: 150, loss: 0.000270505144726485
step: 160, loss: 0.00026580423582345247
step: 170, loss: 0.0002732225111685693
step: 180, loss: 0.004822565242648125
step: 190, loss: 0.0015448404010385275
step: 200, loss: 0.0011830354342237115
step: 210, loss: 0.00029367837123572826
step: 220, loss: 0.00011360996722942218
```

```
step: 230, loss: 0.0013801440363749862
step: 240, loss: 0.00010795117123052478
step: 250, loss: 0.00013317940465640277
step: 260, loss: 9.302348917117342e-05
step: 270, loss: 0.004189082887023687
step: 280, loss: 8.530450577381998e-05
step: 290, loss: 9.530016541248187e-05
step: 300, loss: 0.0001709491480141878
step: 310, loss: 0.002067041350528598
step: 320, loss: 0.0006053319666534662
step: 330, loss: 0.0001077880006050691
step: 340, loss: 5.724294169340283e-05
step: 350, loss: 5.977608452667482e-05
step: 360, loss: 0.002294057048857212
step: 370, loss: 0.004208848811686039
step: 380, loss: 6.535309512401e-05
step: 390, loss: 0.00010674908116925508
step: 400, loss: 0.00021916769037488848
step: 410, loss: 0.00031807785853743553
step: 420, loss: 0.0009258968057110906
step: 430, loss: 0.0003762409614864737
step: 440, loss: 5.147767296875827e-05
step: 450, loss: 9.680978109827265e-05
step: 460, loss: 8.195445843739435e-05
step: 470, loss: 9.100536408368498e-05
step: 480, loss: 0.00011139416164951399
step: 490, loss: 0.06431901454925537
step: 500, loss: 0.00026265691849403083
step: 510, loss: 0.00030311159207485616
step: 520, loss: 0.0001371303224004805
step: 530, loss: 4.3323543650330976e-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=0.9089210649229332, f1=0.9030131826741995,
best_f1=0.9039704524469068
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001672731013968587
step: 10, loss: 0.0005601735319942236
step: 20, loss: 3.571302295313217e-05
step: 30, loss: 0.0002476587542332709
step: 40, loss: 0.0005160451983101666
step: 50, loss: 0.00013860220497008413
step: 60, loss: 5.1479997637216e-05
step: 70, loss: 5.0079921493306756e-05
step: 80, loss: 8.918515959521756e-05
```

```
step: 90, loss: 8.350372809218243e-05
step: 100, loss: 0.0002480149269104004
step: 110, loss: 0.00010334251419408247
step: 120, loss: 0.0001709375937934965
step: 130, loss: 0.00047773082042112947
step: 140, loss: 8.907634037313983e-05
step: 150, loss: 0.0036916187964379787
step: 160, loss: 0.00015062515740282834
step: 170, loss: 0.00022500002523884177
step: 180, loss: 4.8683465138310567e-05
step: 190, loss: 0.0003805985616054386
step: 200, loss: 7.291706424439326e-05
step: 210, loss: 9.013474482344463e-05
step: 220, loss: 8.419118967140093e-05
step: 230, loss: 0.00010325410403311253
step: 240, loss: 0.004771274980157614
step: 250, loss: 0.0003887685015797615
step: 260, loss: 8.98796715773642e-05
step: 270, loss: 0.0023128390312194824
step: 280, loss: 0.00020969372417312115
step: 290, loss: 9.380467236042023e-05
step: 300, loss: 0.017114030197262764
step: 310, loss: 0.003645631717517972
step: 320, loss: 0.05092315375804901
step: 330, loss: 0.00027899249107576907
step: 340, loss: 0.00013589876471087337
step: 350, loss: 0.05383101850748062
step: 360, loss: 0.004540913738310337
step: 370, loss: 4.1385199438082054e-05
step: 380, loss: 4.3806390749523416e-05
step: 390, loss: 0.00020796930766664445
step: 400, loss: 3.191713403793983e-05
step: 410, loss: 2.914600190706551e-05
step: 420, loss: 0.0028313971124589443
step: 430, loss: 0.0007755258120596409
step: 440, loss: 7.157215441111475e-05
step: 450, loss: 3.751185431610793e-05
step: 460, loss: 0.0004198942333459854
step: 470, loss: 0.0007541536469943821
step: 480, loss: 0.0003891197557095438
step: 490, loss: 0.00038096753996796906
step: 500, loss: 0.00021654368902090937
step: 510, loss: 4.430636909091845e-05
step: 520, loss: 4.46745652880054e-05
step: 530, loss: 0.00042977489647455513
/usr/local/lib/python3.7/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=0.9083215796897038, f1=0.8959318826868495,
best_f1=0.9039704524469068
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.944512279005721e-05
step: 10, loss: 0.00018387407180853188
step: 20, loss: 2.754412889771629e-05
step: 30, loss: 5.0363400077912956e-05
step: 40, loss: 0.00015681402874179184
step: 50, loss: 5.36187544639688e-05
step: 60, loss: 0.0008699352038092911
step: 70, loss: 0.005944648291915655
step: 80, loss: 0.00032148530590347946
step: 90, loss: 0.012884755618870258
step: 100, loss: 8.511572377756238e-05
step: 110, loss: 3.261018719058484e-05
step: 120, loss: 0.0011640549637377262
step: 130, loss: 0.0002504217845853418
step: 140, loss: 0.0001843659410951659
step: 150, loss: 0.0004465110250748694
step: 160, loss: 6.995621515670791e-05
step: 170, loss: 0.019902531057596207
step: 180, loss: 6.893208774272352e-05
step: 190, loss: 4.152579276706092e-05
step: 200, loss: 6.865258183097467e-05
step: 210, loss: 0.00019546839757822454
step: 220, loss: 8.925353176891804e-05
step: 230, loss: 7.85619777161628e-05
step: 240, loss: 4.208603786537424e-05
step: 250, loss: 0.001235414994880557
step: 260, loss: 0.004441940691322088
step: 270, loss: 8.126660395646468e-05
step: 280, loss: 0.0007078108610585332
step: 290, loss: 0.005731785204261541
step: 300, loss: 6.189584382809699e-05
step: 310, loss: 6.0706650401698425e-05
step: 320, loss: 5.5405296734534204e-05
step: 330, loss: 4.1382259951205924e-05
step: 340, loss: 6.989809480728582e-05
step: 350, loss: 0.005777871701866388
step: 360, loss: 0.00020107523596379906
step: 370, loss: 0.0002880293468479067
step: 380, loss: 0.00037341416464187205
step: 390, loss: 0.0001986924180528149
step: 400, loss: 4.314757097745314e-05
step: 410, loss: 6.643805681960657e-05
step: 420, loss: 0.00012297966168262064
```

```
step: 430, loss: 6.6427506681066e-05
step: 440, loss: 5.7708977692527696e-05
step: 450, loss: 0.00011159385030623525
step: 460, loss: 6.621429201913998e-05
step: 470, loss: 0.00010685995948733762
step: 480, loss: 5.720036642742343e-05
step: 490, loss: 0.00037849892396479845
step: 500, loss: 0.0010735387913882732
step: 510, loss: 5.3708459745394066e-05
step: 520, loss: 0.0002688208478502929
step: 530, loss: 0.000652708753477782
/usr/local/lib/python3.7/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=0.9051643192488262, f1=0.89413988657845,
best_f1=0.9039704524469068
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.695786578347906e-05
step: 10, loss: 0.0033656007144600153
step: 20, loss: 0.0015607330715283751
step: 30, loss: 0.001548310974612832
step: 40, loss: 0.00015376319061033428
step: 50, loss: 8.842317765811458e-05
step: 60, loss: 3.68084802175872e-05
step: 70, loss: 0.0011569138150662184
step: 80, loss: 0.00018122588517144322
step: 90, loss: 0.00011651258682832122
step: 100, loss: 3.7141242501093075e-05
step: 110, loss: 0.00024959215079434216
step: 120, loss: 0.0014166616601869464
step: 130, loss: 7.06295613781549e-05
step: 140, loss: 5.7254703278886154e-05
step: 150, loss: 7.136238127714023e-05
step: 160, loss: 5.0770682719303295e-05
step: 170, loss: 6.292772013694048e-05
step: 180, loss: 3.3921536669367924e-05
step: 190, loss: 5.366226832848042e-05
step: 200, loss: 0.001977220643311739
step: 210, loss: 0.0019514269661158323
step: 220, loss: 4.439812619239092e-05
step: 230, loss: 2.4641974960104562e-05
step: 240, loss: 0.00019227739539928734
step: 250, loss: 0.00079876504605636
step: 260, loss: 3.500135426293127e-05
step: 270, loss: 4.4684511522063985e-05
step: 280, loss: 0.0005210753879509866
```

```
step: 290, loss: 0.00017952141934074461
step: 300, loss: 0.0004881895729340613
step: 310, loss: 2.073841096716933e-05
step: 320, loss: 5.1107661420246586e-05
step: 330, loss: 0.004188857972621918
step: 340, loss: 4.8335317842429504e-05
step: 350, loss: 0.0003189364215359092
step: 360, loss: 4.317050479585305e-05
step: 370, loss: 0.0007825332577340305
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 380, loss: 7.754963007755578e-05
step: 390, loss: 7.185182766988873e-05
step: 400, loss: 3.6799207009607926e-05
step: 410, loss: 3.649117934401147e-05
step: 420, loss: 0.0007405329961329699
step: 430, loss: 3.8382462662411854e-05
step: 440, loss: 0.00016950065037235618
step: 450, loss: 7.819433085387573e-05
step: 460, loss: 9.18054865906015e-05
step: 470, loss: 0.004276062827557325
step: 480, loss: 4.080418511875905e-05
step: 490, loss: 4.0562081267125905e-05
step: 500, loss: 8.8113134552259e-05
step: 510, loss: 3.203668529749848e-05
step: 520, loss: 0.00043465339695103467
step: 530, loss: 4.0353981603402644e-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=0.9029815428300995, f1=0.9016627078384797,
best_f1=0.9039704524469068
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00015857847756706178
step: 10, loss: 7.979221118148416e-05
step: 20, loss: 2.4664590455358848e-05
step: 30, loss: 0.0003265290579292923
step: 40, loss: 3.6445642763283104e-05
step: 50, loss: 0.0005233414121903479
step: 60, loss: 3.0125445846351795e-05
step: 70, loss: 0.000620041333604604
step: 80, loss: 0.000243779577431269
step: 90, loss: 4.6231827582232654e-05
step: 100, loss: 4.764397090184502e-05
step: 110, loss: 7.579204248031601e-05
step: 120, loss: 4.315558908274397e-05
step: 130, loss: 4.4565240386873484e-05
```

```
step: 140, loss: 0.01811087876558304
step: 150, loss: 0.0005344083183445036
step: 160, loss: 0.0004301058070268482
step: 170, loss: 5.43368405487854e-05
step: 180, loss: 0.00019427179358899593
step: 190, loss: 0.07796333730220795
step: 200, loss: 0.0003624391392804682
step: 210, loss: 0.0007065151003189385
step: 220, loss: 0.00020902618416585028
step: 230, loss: 0.00011789756536018103
step: 240, loss: 6.836531247245148e-05
step: 250, loss: 4.6979064791230485e-05
step: 260, loss: 2.5595811166567728e-05
step: 270, loss: 0.007623456884175539
step: 280, loss: 3.9146612834883854e-05
step: 290, loss: 0.00015247586998157203
step: 300, loss: 3.312737317173742e-05
step: 310, loss: 2.937653880508151e-05
step: 320, loss: 0.002830727491527796
step: 330, loss: 4.1444240196142346e-05
step: 340, loss: 4.074570824741386e-05
step: 350, loss: 8.425895066466182e-05
step: 360, loss: 0.0002657863951753825
step: 370, loss: 0.00025721167912706733
step: 380, loss: 4.493632513913326e-05
step: 390, loss: 0.0001988460135180503
step: 400, loss: 7.04187696101144e-05
step: 410, loss: 0.0002453791385050863
step: 420, loss: 4.977460775990039e-05
step: 430, loss: 0.0002045117726083845
step: 440, loss: 1.79034577740822e-05
step: 450, loss: 0.00021924084285274148
step: 460, loss: 6.818126712460071e-05
step: 470, loss: 0.0035196540411561728
step: 480, loss: 3.1373212550533935e-05
step: 490, loss: 4.775110573973507e-05
step: 500, loss: 0.00032173143699765205
step: 510, loss: 4.017445826320909e-05
step: 520, loss: 5.6062683142954484e-05
step: 530, loss: 0.00018331459432374686
/usr/local/lib/python3.7/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=0.9036144578313254, f1=0.9074848907484891,
best_f1=0.9039704524469068
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 4.426408122526482e-05
step: 10, loss: 4.040641215397045e-05
step: 20, loss: 5.7976634707301855e-05
step: 30, loss: 3.942261901102029e-05
step: 40, loss: 2.745462188613601e-05
step: 50, loss: 2.1945070329820737e-05
step: 60, loss: 5.662025068886578e-05
step: 70, loss: 3.61925813319483e-05
step: 80, loss: 8.914498903322965e-05
step: 90, loss: 9.890960063785315e-05
step: 100, loss: 2.7525422410690226e-05
step: 110, loss: 0.0005083226133137941
step: 120, loss: 0.0010827264050021768
step: 130, loss: 0.00022481533233076334
step: 140, loss: 4.166801954852417e-05
step: 150, loss: 4.93105108034797e-05
step: 160, loss: 3.1559146009385586e-05
step: 170, loss: 0.00014290487160906196
step: 180, loss: 2.113686059601605e-05
step: 190, loss: 7.189540338004008e-05
step: 200, loss: 2.9693379474338144e-05
step: 210, loss: 7.973132596816868e-05
step: 220, loss: 2.1293310055625625e-05
step: 230, loss: 0.0001307603670284152
step: 240, loss: 5.1759270718321204e-05
step: 250, loss: 0.0019581655506044626
step: 260, loss: 9.97993047349155e-05
step: 270, loss: 0.00020203769963700324
step: 280, loss: 5.9923699154751375e-05
step: 290, loss: 0.00011090880434494466
step: 300, loss: 0.0053869495168328285
step: 310, loss: 7.061992801027372e-05
step: 320, loss: 3.480057785054669e-05
step: 330, loss: 0.003543006954714656
step: 340, loss: 8.996052929433063e-05
step: 350, loss: 0.036177389323711395
step: 360, loss: 3.156342791044153e-05
step: 370, loss: 2.5774452296900563e-05
step: 380, loss: 0.0001486000546719879
step: 390, loss: 4.925384928355925e-05
step: 400, loss: 0.000313704862492159
step: 410, loss: 3.8076905184425414e-05
step: 420, loss: 7.300591823877767e-05
step: 430, loss: 3.424124588491395e-05
step: 440, loss: 9.998495079344139e-05
step: 450, loss: 0.001186381559818983
step: 460, loss: 5.891470209462568e-05
step: 470, loss: 7.3326300480403e-05
```

```
step: 480, loss: 2.516362837923225e-05
step: 490, loss: 0.0037618866190314293
step: 500, loss: 5.947229146840982e-05
step: 510, loss: 0.01930922269821167
step: 520, loss: 8.649734809296206e-05
step: 530, loss: 3.173109871568158e-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=0.9053613053613055, f1=0.9097709209911174,
best_f1=0.9039704524469068
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.")
5742it [00:15, 362.62it/s]
load_f1 = 0.9141531322505799
real_f1 = 0.9163568773234201
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, 386.44it/s]
```

3 DK

```
[ ]:  # Para executar 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: 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: 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: 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: 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: 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: 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: 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: 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: 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: 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: 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: 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: 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: 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: 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=170b4f36d0dd08001db0f4d0eec58d661d8c85b02dc08264cb132953605d7a4f
  Stored in directory: /tmp/pip-ephem-wheel-cache-o0zfgssc/wheels/11/95/ba/2c36c
c368c0bd339b44a791c2c1881a1fb714b78c29a4cb8f5
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 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)…

245

```
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: 0.8769121170043945
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=0.358974358974359, f1=0.3146067415730337,
best_f1=0.3146067415730337
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.35865864157676697
/usr/local/lib/python3.7/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=0.2711864406779661, f1=0.2816901408450704,
best_f1=0.3146067415730337
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3311608135700226
```

```
/usr/local/lib/python3.7/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=0.2692307692307693, f1=0.2666666666666667,
best_f1=0.3146067415730337
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.37242746353149414
/usr/local/lib/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=0.4210526315789474, f1=0.3013698630136986,
best_f1=0.3013698630136986
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.29269978404045105
/usr/local/lib/python3.7/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=0.4126984126984127, f1=0.3373493975903615,
best_f1=0.3013698630136986
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2748320698738098
/usr/local/lib/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=0.45283018867924535, 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: 0.2668851315975189
/usr/local/lib/python3.7/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=0.42553191489361697, f1=0.38596491228070184,
```

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: 0.41150376200675964
/usr/local/lib/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=0.5, f1=0.4482758620689656, best_f1=0.4482758620689656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.21257396042346954
/usr/local/lib/python3.7/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=0.43137254901960786, f1=0.42857142857142855,
best_f1=0.4482758620689656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.29201075434684753
/usr/local/lib/python3.7/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=0.43902439024390244, f1=0.48, best_f1=0.4482758620689656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2737025320529938
/usr/local/lib/python3.7/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=0.5, f1=0.4897959183673469, best_f1=0.4482758620689656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.25304538011550903
/usr/local/lib/python3.7/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=0.4736842105263159, f1=0.45833333333333326,
best_f1=0.4482758620689656
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.17561988532543182
/usr/local/lib/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=0.5116279069767441, f1=0.45833333333333326,
best_f1=0.45833333333333326
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1889357566833496
/usr/local/lib/python3.7/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=0.5116279069767441, f1=0.45833333333333326,
best_f1=0.45833333333333326
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.27296966314315796
/usr/local/lib/python3.7/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=0.5116279069767441, f1=0.45833333333333326,
best_f1=0.45833333333333326
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, 126636.25it/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 = 0.4761904761904762
real_f1 = 0.4761904761904762
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 [00:16, 268.86it/s]
```

3.1.2 DBLP-ACM - Running the matcher

```python
[ ]: # 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 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: 0.7971441745758057
step: 10, loss: 0.48138856887817383
step: 20, loss: 0.6127474308013916

251
```

```
step: 30, loss: 0.5085351467132568
step: 40, loss: 0.42998477816581726
step: 50, loss: 0.3414802849292755
step: 60, loss: 0.24683596193790436
step: 70, loss: 0.1364862322807312
step: 80, loss: 0.12910018861293793
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 90, loss: 0.021884815767407417
step: 100, loss: 0.16252119839191437
step: 110, loss: 0.04280352219939232
step: 120, loss: 0.0520084947347641
step: 130, loss: 0.13457149267196655
step: 140, loss: 0.18062494695186615
step: 150, loss: 0.13240154087543488
step: 160, loss: 0.22006531059741974
step: 170, loss: 0.0069502610713243484
step: 180, loss: 0.011754951439797878
step: 190, loss: 0.11716275662183762
step: 200, loss: 0.01257830299437046
step: 210, loss: 0.014047220349311829
step: 220, loss: 0.00679743429645896
step: 230, loss: 0.005398761946707964
/usr/local/lib/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=0.970917225950783, f1=0.9570135746606334,
best_f1=0.9570135746606334
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06442093104124069
step: 10, loss: 0.0019813545513898134
step: 20, loss: 0.010282019153237343
step: 30, loss: 0.007297302596271038
step: 40, loss: 0.0272186491638422
step: 50, loss: 0.010458247736096382
step: 60, loss: 0.08646419644355774
step: 70, loss: 0.059675183147192
step: 80, loss: 0.0070898146368563175
step: 90, loss: 0.004649091511964798
step: 100, loss: 0.19660773873329163
step: 110, loss: 0.08658973127603531
step: 120, loss: 0.017047224566340446
```

```
step: 130, loss: 0.07994914054870605
step: 140, loss: 0.10261448472738266
step: 150, loss: 0.006054380442947149
step: 160, loss: 0.029169779270887375
step: 170, loss: 0.07379808276891708
step: 180, loss: 0.0026523214764893055
step: 190, loss: 0.06895451247692108
step: 200, loss: 0.006674130912870169
step: 210, loss: 0.12359055131673813
step: 220, loss: 0.0009558204910717905
step: 230, loss: 0.09348660707473755
/usr/local/lib/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=0.9743589743589743, f1=0.9598214285714285,
best_f1=0.9598214285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02323567494750023
step: 10, loss: 0.1287080943584442
step: 20, loss: 0.053924594074487686
step: 30, loss: 0.010514660738408566
step: 40, loss: 0.018465286120772362
step: 50, loss: 0.021387146785855293
step: 60, loss: 0.0006444465252570808
step: 70, loss: 0.007132370490580797
step: 80, loss: 0.005730516742914915
step: 90, loss: 0.015413918532431126
step: 100, loss: 0.002817385597154498
step: 110, loss: 0.058364201337099075
step: 120, loss: 0.08242657780647278
step: 130, loss: 0.0008438770310021937
step: 140, loss: 0.020765507593750954
step: 150, loss: 0.0020745513029396534
step: 160, loss: 0.006124912295490503
step: 170, loss: 0.0036801169626414776
step: 180, loss: 0.004141069948673248
step: 190, loss: 0.020739082247018814
step: 200, loss: 0.0015794796636328101
step: 210, loss: 0.04243094101548195
step: 220, loss: 0.0046340785920619965
step: 230, loss: 0.1036260798573494
/usr/local/lib/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=0.9777777777777777, 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: 0.005861867684870958
step: 10, loss: 0.015471928752958775
step: 20, loss: 0.00041279109427705407
step: 30, loss: 0.00021875124366488308
step: 40, loss: 0.0014712696429342031
step: 50, loss: 0.003923339769244194
step: 60, loss: 0.0004888083203695714
step: 70, loss: 0.0015575832221657038
step: 80, loss: 0.022987214848399162
step: 90, loss: 0.22305722534656525
step: 100, loss: 0.005293748341500759
step: 110, loss: 0.008562845177948475
step: 120, loss: 0.08368875831365585
step: 130, loss: 0.0011394221801310778
step: 140, loss: 0.00641384394839406
step: 150, loss: 0.010087274014949799
step: 160, loss: 0.0016500415513291955
step: 170, loss: 0.02095113880932331
step: 180, loss: 0.0736365094780922
step: 190, loss: 0.007358016446232796
step: 200, loss: 0.004841885529458523
step: 210, loss: 0.004411677829921246
step: 220, loss: 0.0005529565969482064
step: 230, loss: 0.0008238345617428422
/usr/local/lib/python3.7/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=0.9690265486725664, f1=0.966740576496674,
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: 0.0016328011406585574
step: 10, loss: 0.035297926515340805
step: 20, loss: 0.010730499401688576
step: 30, loss: 0.002450797241181135
step: 40, loss: 0.014457807876169682
```

```
step: 50, loss: 0.00026490268646739423
step: 60, loss: 0.00029603595612570643
step: 70, loss: 0.0002152677916456014
step: 80, loss: 0.0004515676700975746
step: 90, loss: 0.013866066932678223
step: 100, loss: 0.012165587395429611
step: 110, loss: 0.006062759086489677
step: 120, loss: 0.1570526510477066
step: 130, loss: 0.0008855434716679156
step: 140, loss: 0.00012871941726189107
step: 150, loss: 0.00012370350304991007
step: 160, loss: 0.0019985237158834934
step: 170, loss: 0.024988600984215736
step: 180, loss: 0.0015013021184131503
step: 190, loss: 0.0003267016727477312
step: 200, loss: 0.004675957374274731
step: 210, loss: 0.0012589386897161603
step: 220, loss: 0.003167507704347372
step: 230, loss: 0.0031409948132932186
/usr/local/lib/python3.7/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=0.9663677130044843, f1=0.9627959413754228,
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: 0.0013640100369229913
step: 10, loss: 0.007041348144412041
step: 20, loss: 0.0106064323335886
step: 30, loss: 0.00028705113800242543
step: 40, loss: 0.0005486909649334848
step: 50, loss: 0.03651507943868637
step: 60, loss: 0.0008057118975557387
step: 70, loss: 0.0004475804162211716
step: 80, loss: 0.002028119983151555
step: 90, loss: 0.0002219670859631151
step: 100, loss: 0.00021825020667165518
step: 110, loss: 0.03430471941828728
step: 120, loss: 0.0027059230487793684
step: 130, loss: 0.0082743801176548
step: 140, loss: 0.006752931512892246
step: 150, loss: 0.0010423630010336637
step: 160, loss: 0.053576067090034485
step: 170, loss: 0.0006005655159242451
step: 180, loss: 0.0007219088729470968
step: 190, loss: 0.0003270096785854548
step: 200, loss: 0.004617572762072086
```

```
step: 210, loss: 0.002853484358638525
step: 220, loss: 0.00029770241235382855
step: 230, loss: 0.09675063192844391
/usr/local/lib/python3.7/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=0.9707207207207207, f1=0.9694224235560589,
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: 0.007204663008451462
step: 10, loss: 0.0002442534314468503
step: 20, loss: 0.0003460158477537334
step: 30, loss: 0.00036878001992590725
step: 40, loss: 0.0018304003169760108
step: 50, loss: 0.00011984690354438499
step: 60, loss: 0.00038463942473754287
step: 70, loss: 0.23165592551231384
step: 80, loss: 0.00016032863641157746
step: 90, loss: 0.0009045269689522684
step: 100, loss: 0.008761064149439335
step: 110, loss: 0.00384707422927022
step: 120, loss: 0.00033137318678200245
step: 130, loss: 0.00025447746156714857
step: 140, loss: 0.00017935257346834987
step: 150, loss: 0.00016149523435160518
step: 160, loss: 0.00035424393718130887
step: 170, loss: 0.00014840118819847703
step: 180, loss: 0.0004458134644664824
step: 190, loss: 0.0003419704153202474
step: 200, loss: 0.0007891985587775707
step: 210, loss: 0.00041461453656666598
step: 220, loss: 0.0027915108948946
step: 230, loss: 0.008260932750999928
/usr/local/lib/python3.7/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=0.9673790776152981, f1=0.9662162162162162,
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: 0.033885687589645386
step: 10, loss: 0.0025333198718726635
step: 20, loss: 0.0005219669546931982
step: 30, loss: 0.0011188397184014328
step: 40, loss: 0.00023021311790216714
```

```
step: 50, loss: 0.00018492281378712505
step: 60, loss: 0.00016985964612104
step: 70, loss: 0.00014971022028476
step: 80, loss: 0.00027035779203288257
step: 90, loss: 0.00021234640735201538
step: 100, loss: 0.000370849302271381
step: 110, loss: 0.00014817748160567135
step: 120, loss: 0.00010438740719109774
step: 130, loss: 0.0002547878830228001
step: 140, loss: 0.0021183881908655167
step: 150, loss: 9.619628690415993e-05
step: 160, loss: 0.00014563914737664163
step: 170, loss: 0.00011046520376112312
step: 180, loss: 0.0005017481162212789
step: 190, loss: 9.456843690713868e-05
step: 200, loss: 0.00015022441220935434
step: 210, loss: 6.35032047284767e-05
step: 220, loss: 0.00033345489646308124
step: 230, loss: 0.009278159588575363
/usr/local/lib/python3.7/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=0.9749430523917996, f1=0.9694224235560589,
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: 6.339112587738782e-05
step: 10, loss: 0.00011548707698239014
step: 20, loss: 0.0011350598651915789
step: 30, loss: 0.0002062375860987231
step: 40, loss: 0.00036344697582535446
step: 50, loss: 0.00014876325440127403
step: 60, loss: 0.0001509926514700055
step: 70, loss: 0.0005054495995864272
step: 80, loss: 0.12397728860378265
step: 90, loss: 0.005629525985568762
step: 100, loss: 0.0011461629765108228
step: 110, loss: 0.0014878390356898308
step: 120, loss: 0.0020079368259757757
step: 130, loss: 0.00015326014545280486
step: 140, loss: 0.000332571828039363
step: 150, loss: 0.00011800610081991181
step: 160, loss: 0.0035427636466920376
step: 170, loss: 0.00024603575002402067
step: 180, loss: 0.0009110851096920669
step: 190, loss: 0.00011809143325081095
step: 200, loss: 0.00019450434774626046
```

```
step: 210, loss: 0.0010620774701237679
step: 220, loss: 0.07437395304441452
step: 230, loss: 0.0053456025198102
/usr/local/lib/python3.7/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=0.9729119638826186, f1=0.9683972911963882,
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: 0.00017196565750055015
step: 10, loss: 0.00010959716746583581
step: 20, loss: 8.882291876943782e-05
step: 30, loss: 0.00016049780242610723
step: 40, loss: 0.00014560170529875904
step: 50, loss: 0.0005505103035829961
step: 60, loss: 0.00014989773626439273
step: 70, loss: 0.0007119251531548798
step: 80, loss: 0.0003016425180248916
step: 90, loss: 0.0005016293143853545
step: 100, loss: 0.00012044620234519243
step: 110, loss: 0.00024549508816562593
step: 120, loss: 0.004999023862183094
step: 130, loss: 0.00023731740657240152
step: 140, loss: 0.016076337546110153
step: 150, loss: 0.00014013804320711643
step: 160, loss: 0.0002004855195991695
step: 170, loss: 0.00011457655637059361
step: 180, loss: 0.00012027277261950076
step: 190, loss: 0.0004771137028001249
step: 200, loss: 0.00012035483086947352
step: 210, loss: 9.166279778582975e-05
step: 220, loss: 0.0007384862983599305
step: 230, loss: 9.654263703851029e-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=0.9696287964004499, f1=0.9695603156708005,
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: 0.00017303841013927013
step: 10, loss: 6.145863881101832e-05
step: 20, loss: 8.55574689921923e-05
step: 30, loss: 0.00021205835219006985
step: 40, loss: 0.002806599484756589
```

```
step: 50, loss: 0.0003848798805847764
step: 60, loss: 0.00016469552065245807
step: 70, loss: 0.0013719910057261586
step: 80, loss: 0.0010340956505388021
step: 90, loss: 5.607675848295912e-05
step: 100, loss: 0.0001118256404879503
step: 110, loss: 5.5271986639127135e-05
step: 120, loss: 8.048495510593057e-05
step: 130, loss: 4.807556251762435e-05
step: 140, loss: 8.370748400921002e-05
step: 150, loss: 5.015380884287879e-05
step: 160, loss: 7.587954314658418e-05
step: 170, loss: 0.03381659835577011
step: 180, loss: 0.00012175336451036856
step: 190, loss: 0.0025360521394759417
step: 200, loss: 7.27379010641016e-05
step: 210, loss: 5.121623325976543e-05
step: 220, loss: 7.520416693296283e-05
step: 230, loss: 0.02369583770632744
/usr/local/lib/python3.7/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=0.9741282339707535, f1=0.9717514124293786,
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: 0.00040400284342467785
step: 10, loss: 0.00033604662166908383
step: 20, loss: 0.00013688605395145714
step: 30, loss: 8.495297515764832e-05
step: 40, loss: 0.00015863279986660928
step: 50, loss: 0.000738665577955544
step: 60, loss: 0.0005578828859142959
step: 70, loss: 7.330378866754472e-05
step: 80, loss: 8.161977166309953e-05
step: 90, loss: 5.212449104874395e-05
step: 100, loss: 4.240695125190541e-05
step: 110, loss: 5.2452956879278645e-05
step: 120, loss: 0.0001246215688297525
step: 130, loss: 0.0003388676850590855
step: 140, loss: 4.654880831367336e-05
step: 150, loss: 4.2126546759391204e-05
step: 160, loss: 0.00048609322402626276
step: 170, loss: 6.45276450086385e-05
step: 180, loss: 6.884251342853531e-05
step: 190, loss: 5.0858285248978063e-05
step: 200, loss: 0.0006844309973530471
```

```
step: 210, loss: 0.013513928279280663
step: 220, loss: 0.00010330423538107425
step: 230, loss: 0.000714607653208077
/usr/local/lib/python3.7/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=0.9700332963374029, f1=0.9730941704035874,
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: 0.00012528487422969192
step: 10, loss: 0.00036162871401757
step: 20, loss: 0.00027251988649368286
step: 30, loss: 0.0026424082461744547
step: 40, loss: 0.0017556938109919429
step: 50, loss: 6.750061584170908e-05
step: 60, loss: 5.393187529989518e-05
step: 70, loss: 6.024767571943812e-05
step: 80, loss: 0.00017635576659813523
step: 90, loss: 0.0001159084786195308
step: 100, loss: 6.605970702366903e-05
step: 110, loss: 6.068783477530815e-05
step: 120, loss: 7.17723451089114e-05
step: 130, loss: 0.00016338244313374162
step: 140, loss: 8.398358477279544e-05
step: 150, loss: 0.00044213375076651573
step: 160, loss: 7.438616739818826e-05
step: 170, loss: 5.332435466698371e-05
step: 180, loss: 0.0001656525128055364
step: 190, loss: 4.6470209781546146e-05
step: 200, loss: 8.955226803664118e-05
step: 210, loss: 0.0012237665941938758
step: 220, loss: 6.279633817030117e-05
step: 230, loss: 5.073791180620901e-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=0.9707207207207207, f1=0.9705882352941176,
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: 9.705561387818307e-05
step: 10, loss: 9.492552635492757e-05
step: 20, loss: 0.0011086120503023267
step: 30, loss: 0.0002280273038195446
step: 40, loss: 3.8844944356242195e-05
```

```
step: 50, loss: 8.278049062937498e-05
step: 60, loss: 4.9094636779045686e-05
step: 70, loss: 5.0445036322344095e-05
step: 80, loss: 8.654611883684993e-05
step: 90, loss: 7.090003782650456e-05
step: 100, loss: 0.001603966928087175
step: 110, loss: 0.0003670515725389123
step: 120, loss: 0.00020919143571518362
step: 130, loss: 0.00010667595051927492
step: 140, loss: 6.589750410057604e-05
step: 150, loss: 0.00010157637007068843
step: 160, loss: 3.6100005672778934e-05
step: 170, loss: 0.00014798885968048126
step: 180, loss: 0.00010449118417454883
step: 190, loss: 0.00013727793702855706
step: 200, loss: 4.432155401445925e-05
step: 210, loss: 0.0001806648215278983
step: 220, loss: 6.836403190391138e-05
step: 230, loss: 2.817378845065832e-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=0.9719416386083053, f1=0.971815107102593,
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: 4.035756865050644e-05
step: 10, loss: 0.00023126068117562681
step: 20, loss: 0.00012023797899018973
step: 30, loss: 8.500088733853772e-05
step: 40, loss: 0.0001009302432066761
step: 50, loss: 6.661035877186805e-05
step: 60, loss: 0.000195499203982763
step: 70, loss: 7.494421151932329e-05
step: 80, loss: 5.154040627530776e-05
step: 90, loss: 3.356339584570378e-05
step: 100, loss: 0.00012063542089890689
step: 110, loss: 6.794734508730471e-05
step: 120, loss: 0.13564926385879517
step: 130, loss: 0.0001001140044536416
step: 140, loss: 0.0001263044832739979
step: 150, loss: 0.00025977680343203247
step: 160, loss: 4.120731318835169e-05
step: 170, loss: 4.9993286665994674e-05
step: 180, loss: 7.237595855258405e-05
step: 190, loss: 0.0003825988678727299
step: 200, loss: 8.513433567713946e-05
```

```
step: 210, loss: 0.006949456874281168
step: 220, loss: 5.416920976131223e-05
step: 230, loss: 8.269414684036747e-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=0.9720044792833147, f1=0.972972972972973,
best_f1=0.9688195991091313
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, 236.25it/s]
load_f1 = 0.9765363128491621
real_f1 = 0.9754464285714286
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 [00:16, 260.03it/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 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: 0.7924099564552307
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.46124163269996643
step: 20, loss: 0.5231547355651855
step: 30, loss: 0.4438588321208954
step: 40, loss: 0.4883056581020355
step: 50, loss: 0.4330879747867584
step: 60, loss: 0.44488510489463806
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 70, loss: 0.19834551215171814
step: 80, loss: 0.18686515092849731
step: 90, loss: 0.1377580463886261
step: 100, loss: 0.3265272378921509
step: 110, loss: 0.19482457637786865
step: 120, loss: 0.14257754385471344
step: 130, loss: 0.05472497269511223
step: 140, loss: 0.279985249042511
step: 150, loss: 0.12542128562927246
step: 160, loss: 0.338727742433548
step: 170, loss: 0.4379519820213318
step: 180, loss: 0.14841340482234955
step: 190, loss: 0.1617499440908432
step: 200, loss: 0.10757837444543839
step: 210, loss: 0.10067395120859146
step: 220, loss: 0.14214082062244415
```

```
step: 230, loss: 0.10811727494001389
step: 240, loss: 0.17133483290672302
step: 250, loss: 0.07953298091888428
step: 260, loss: 0.0413941890001297
step: 270, loss: 0.11792970448732376
step: 280, loss: 0.13387079536914825
step: 290, loss: 0.11840201169252396
step: 300, loss: 0.13221246004104614
step: 310, loss: 0.16970475018024445
step: 320, loss: 0.16635887324810028
step: 330, loss: 0.19747692346572876
step: 340, loss: 0.15303297340869904
step: 350, loss: 0.14308585226535797
step: 360, loss: 0.09377992898225784
step: 370, loss: 0.16836795210838318
step: 380, loss: 0.13831539452075958
step: 390, loss: 0.11156544089317322
step: 400, loss: 0.01677999086678028
step: 410, loss: 0.11242063343524933
step: 420, loss: 0.03436867892742157
step: 430, loss: 0.11629840731620789
step: 440, loss: 0.23745699226856232
step: 450, loss: 0.018993055447936058
step: 460, loss: 0.1127573847706909
step: 470, loss: 0.2523824870586395
step: 480, loss: 0.2635113000869751
step: 490, loss: 0.08883830904960632
step: 500, loss: 0.0277822595089674
step: 510, loss: 0.10922891646623611
step: 520, loss: 0.08884138613939285
step: 530, loss: 0.2635475695133209
/usr/local/lib/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=0.8920056100981766, f1=0.8902382064455862,
best_f1=0.8902382064455862
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2386055290699005
step: 10, loss: 0.11600880324840546
step: 20, loss: 0.15492823719978333
step: 30, loss: 0.13924865424633026
step: 40, loss: 0.023804865777492523
```

```
step: 50, loss: 0.10340335220098495
step: 60, loss: 0.17240574955940247
step: 70, loss: 0.34243106842041016
step: 80, loss: 0.02026698924601078
step: 90, loss: 0.09335444867610931
step: 100, loss: 0.30028337240219116
step: 110, loss: 0.07515314966440201
step: 120, loss: 0.11114243417978287
step: 130, loss: 0.08118550479412079
step: 140, loss: 0.037513043731451035
step: 150, loss: 0.03078765980899334
step: 160, loss: 0.15260060131549835
step: 170, loss: 0.04737955331802368
step: 180, loss: 0.018871331587433815
step: 190, loss: 0.0924186259508133
step: 200, loss: 0.049342140555381775
step: 210, loss: 0.1017257422208786
step: 220, loss: 0.1053803488612175
step: 230, loss: 0.09622996300458908
step: 240, loss: 0.1262613832950592
step: 250, loss: 0.04965320974588394
step: 260, loss: 0.042372044175863266
step: 270, loss: 0.07876594364643097
step: 280, loss: 0.07916798442602158
step: 290, loss: 0.041601214557886124
step: 300, loss: 0.05776198208332062
step: 310, loss: 0.03119998797774315
step: 320, loss: 0.1108069196343422
step: 330, loss: 0.022392740473151207
step: 340, loss: 0.015024282969534397
step: 350, loss: 0.059788625687360764
step: 360, loss: 0.11621053516864777
step: 370, loss: 0.04149946942925453
step: 380, loss: 0.08435884118080139
step: 390, loss: 0.06995288282632828
step: 400, loss: 0.07652588933706284
step: 410, loss: 0.009474245831370354
step: 420, loss: 0.03189869597554207
step: 430, loss: 0.05570197105407715
step: 440, loss: 0.015284432098269463
step: 450, loss: 0.17301814258098602
step: 460, loss: 0.21228013932704926
step: 470, loss: 0.12101995944976807
step: 480, loss: 0.24344760179519653
step: 490, loss: 0.12697802484035492
step: 500, loss: 0.044768452644348145
step: 510, loss: 0.1309777945280075
step: 520, loss: 0.0317250341769867
```

```
step: 530, loss: 0.19641993939876556
/usr/local/lib/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=0.9070422535211268, f1=0.9104477611940298,
best_f1=0.9104477611940298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.039283789694309235
step: 10, loss: 0.12061251699924469
step: 20, loss: 0.04536302387714386
step: 30, loss: 0.25900885462760925
step: 40, loss: 0.01655598171055317
step: 50, loss: 0.012724379077553749
step: 60, loss: 0.021636370569467545
step: 70, loss: 0.0074782841838896275
step: 80, loss: 0.028658907860517502
step: 90, loss: 0.14060568809509277
step: 100, loss: 0.010174209251999855
step: 110, loss: 0.03279908001422882
step: 120, loss: 0.02558637410402298
step: 130, loss: 0.00847102701663971
step: 140, loss: 0.10102445632219315
step: 150, loss: 0.025976212695240974
step: 160, loss: 0.0038301709573715925
step: 170, loss: 0.0491856187582016
step: 180, loss: 0.11022146791219711
step: 190, loss: 0.007169898133724928
step: 200, loss: 0.21952049434185028
step: 210, loss: 0.06911540031433105
step: 220, loss: 0.1312745213508606
step: 230, loss: 0.02332480624318123
step: 240, loss: 0.012308722361922264
step: 250, loss: 0.046046555042266846
step: 260, loss: 0.006014632061123848
step: 270, loss: 0.014142638072371483
step: 280, loss: 0.10293428599834442
step: 290, loss: 0.24774914979934692
step: 300, loss: 0.03833485394716263
step: 310, loss: 0.15181171894073486
step: 320, loss: 0.013337945565581322
step: 330, loss: 0.04627753421664238
step: 340, loss: 0.0174228698015213
```

```
step: 350, loss: 0.13698966801166534
step: 360, loss: 0.011210797354578972
step: 370, loss: 0.021506113931536674
step: 380, loss: 0.07144390046596527
step: 390, loss: 0.071905933320522231
step: 400, loss: 0.04086451232433319
step: 410, loss: 0.08002157509326935
step: 420, loss: 0.075513266602697372
step: 430, loss: 0.04114105552434921
step: 440, loss: 0.05853309482336044
step: 450, loss: 0.15678220987319946
step: 460, loss: 0.13109996914863586
step: 470, loss: 0.009140793234109879
step: 480, loss: 0.05136948078870773
step: 490, loss: 0.020129462704062462
step: 500, loss: 0.1049535870552063
step: 510, loss: 0.018447332084178925
step: 520, loss: 0.008360697887837887
step: 530, loss: 0.01095216628164053
/usr/local/lib/python3.7/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=0.9007285974499089, f1=0.9004092769440655,
best_f1=0.9104477611940298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006792099680751562
step: 10, loss: 0.007050659041851759
step: 20, loss: 0.09040488302707672
step: 30, loss: 0.015588432550430298
step: 40, loss: 0.0059374067932367325
step: 50, loss: 0.046628884971141815
step: 60, loss: 0.0025498028844594955
step: 70, loss: 0.013851379975676537
step: 80, loss: 0.017525173723697662
step: 90, loss: 0.15853913128376007
step: 100, loss: 0.1817716509103775
step: 110, loss: 0.032825812697410583
step: 120, loss: 0.007713764905929565
step: 130, loss: 0.03247793763875961
step: 140, loss: 0.022262969985604286
step: 150, loss: 0.03382158279418945
step: 160, loss: 0.00291619636118412
step: 170, loss: 0.011065076105296612
step: 180, loss: 0.018620261922478676
step: 190, loss: 0.056306954473257065
step: 200, loss: 0.0026280307210981846
```

```
step: 210, loss: 0.04850895330309868
step: 220, loss: 0.12409475445747375
step: 230, loss: 0.4909496009349823
step: 240, loss: 0.00778992148116231
step: 250, loss: 0.018447065725922585
step: 260, loss: 0.12538078427314758
step: 270, loss: 0.14277228713035583
step: 280, loss: 0.004448685795068741
step: 290, loss: 0.03933870419859886
step: 300, loss: 0.015632309019565582
step: 310, loss: 0.0038525047712028027
step: 320, loss: 0.08306685090065002
step: 330, loss: 0.04955463856458664
step: 340, loss: 0.05836660787463188
step: 350, loss: 0.010832889936864376
step: 360, loss: 0.03428703919053078
step: 370, loss: 0.023519566282629967
step: 380, loss: 0.008495645597577095
step: 390, loss: 0.017530502751469612
step: 400, loss: 0.00886064488440752
step: 410, loss: 0.08011895418167114
step: 420, loss: 0.07754646986722946
step: 430, loss: 0.024317583069205284
step: 440, loss: 0.1227385401725769
step: 450, loss: 0.018870238214731216
step: 460, loss: 0.0018563823541626334
step: 470, loss: 0.014403345994651318
step: 480, loss: 0.00887301191687584
step: 490, loss: 0.09100735932588577
step: 500, loss: 0.004918817430734634
step: 510, loss: 0.14019858837127686
step: 520, loss: 0.23201388120651245
step: 530, loss: 0.020666461437940598
/usr/local/lib/python3.7/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=0.9021937842778793, f1=0.897716894977169,
best_f1=0.9104477611940298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006022321060299873
step: 10, loss: 0.06432472169399261
step: 20, loss: 0.047971852123737335
step: 30, loss: 0.013478708453476429
step: 40, loss: 0.04381547123193741
step: 50, loss: 0.003696516389027238
step: 60, loss: 0.0022524225059896708
```

```
step: 70, loss: 0.1324538141489029
step: 80, loss: 0.0017495541833341122
step: 90, loss: 0.005056372378021479
step: 100, loss: 0.0071142567321658134
step: 110, loss: 0.007460595108568668
step: 120, loss: 0.0208884384483099
step: 130, loss: 0.007298387587070465
step: 140, loss: 0.002524443669244647
step: 150, loss: 0.009822946973145008
step: 160, loss: 0.00971466489136219
step: 170, loss: 0.0033675942104309797
step: 180, loss: 0.0014111929340288043
step: 190, loss: 0.017094366252422333
step: 200, loss: 0.01030647661536932
step: 210, loss: 0.007102060131728649
step: 220, loss: 0.0005608184146694839
step: 230, loss: 0.0008852812461555004
step: 240, loss: 0.023265434429049492
step: 250, loss: 0.0012780928518623114
step: 260, loss: 0.03461616113781929
step: 270, loss: 0.05295082554221153
step: 280, loss: 0.07088273763656616
step: 290, loss: 0.0028168438002467155
step: 300, loss: 0.04198722541332245
step: 310, loss: 0.0016504502855241299
step: 320, loss: 0.004131937865167856
step: 330, loss: 0.007252927869558334
step: 340, loss: 0.018119996413588524
step: 350, loss: 0.0027043959125876427
step: 360, loss: 0.1498906910419464
step: 370, loss: 0.0009155795560218394
step: 380, loss: 0.011389797553420067
step: 390, loss: 0.018422577530145645
step: 400, loss: 0.02063353732228279
step: 410, loss: 0.0013614011695608497
step: 420, loss: 0.01434725895524025
step: 430, loss: 0.01264370046555996
step: 440, loss: 0.035095103085041046
step: 450, loss: 0.012083783745765686
step: 460, loss: 0.07523234933614731
step: 470, loss: 0.02508188784122467
step: 480, loss: 0.0019287520553916693
step: 490, loss: 0.0074625094421207905
step: 500, loss: 0.0016763993771746755
step: 510, loss: 0.04651284217834473
step: 520, loss: 0.011798799969255924
step: 530, loss: 0.020899301394820213
/usr/local/lib/python3.7/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=0.9019980970504282, f1=0.8928229665071771,
best_f1=0.9104477611940298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0243715550750494
step: 10, loss: 0.011245988309383392
step: 20, loss: 0.004699685610830784
step: 30, loss: 0.0013107939157634974
step: 40, loss: 0.0007184430141933262
step: 50, loss: 0.013813141733407974
step: 60, loss: 0.03229542076587677
step: 70, loss: 0.01012780237942934
step: 80, loss: 0.0007894731243140996
step: 90, loss: 0.0073320697993040085
step: 100, loss: 0.08287346363067627
step: 110, loss: 0.2106078565120697
step: 120, loss: 0.011587828397750854
step: 130, loss: 0.0012895665131509304
step: 140, loss: 0.004900555592030287
step: 150, loss: 0.001987220486626029
step: 160, loss: 0.003406956559047103
step: 170, loss: 0.0030640934128314257
step: 180, loss: 0.00031467818189412355
step: 190, loss: 0.00764817139133811
step: 200, loss: 0.0002536106330808252
step: 210, loss: 0.0003780378319788724
step: 220, loss: 0.0006502858595922589
step: 230, loss: 0.0029437316115945578
step: 240, loss: 0.00601110327988863
step: 250, loss: 0.0002402377431280911
step: 260, loss: 0.04256713390350342
step: 270, loss: 0.002197777619585395
step: 280, loss: 0.10966670513153076
step: 290, loss: 0.009332867339253426
step: 300, loss: 0.0008973160875029862
step: 310, loss: 0.001181241124868393
step: 320, loss: 0.10431816428899765
step: 330, loss: 0.00404629111289978
step: 340, loss: 0.002001373562961817
step: 350, loss: 0.08453928679227829
step: 360, loss: 0.004227080848067999
step: 370, loss: 0.008362467400729656
step: 380, loss: 0.007225956302136183
step: 390, loss: 0.16290897130966187
step: 400, loss: 0.0003318320959806442
```

```
step: 410, loss: 0.01989017240703106
step: 420, loss: 0.030679097399115562
step: 430, loss: 0.0006147891981527209
step: 440, loss: 0.006477974820882082
step: 450, loss: 0.022982297465205193
step: 460, loss: 0.11566944420337677
step: 470, loss: 0.007529266644269228
step: 480, loss: 0.016544494777917862
step: 490, loss: 0.0006045677582733333
step: 500, loss: 0.004286336712539196
step: 510, loss: 0.0241264458745718
step: 520, loss: 0.006169862579554319
step: 530, loss: 0.002293383702635765
/usr/local/lib/python3.7/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=0.9064272211720227, f1=0.9018867924528302,
best_f1=0.9104477611940298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02899659052491188
step: 10, loss: 0.02447107620537281
step: 20, loss: 0.007908526808023453
step: 30, loss: 0.20208057761192322
step: 40, loss: 0.05590961501002312
step: 50, loss: 0.007333146408200264
step: 60, loss: 0.0005152914091013372
step: 70, loss: 0.0037516988813877106
step: 80, loss: 0.0005301770870573819
step: 90, loss: 0.0005313316360116005
step: 100, loss: 0.0005374237080104649
step: 110, loss: 0.03119746968150139
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 120, loss: 0.00028816916164942086
step: 130, loss: 0.0008188208448700607
step: 140, loss: 0.010508107021450996
step: 150, loss: 0.00035504449624568224
step: 160, loss: 0.0007633817149326205
step: 170, loss: 0.0011460561072453856
step: 180, loss: 0.0005727981915697455
step: 190, loss: 0.0007862074999138713
step: 200, loss: 0.00380427623167634
step: 210, loss: 0.0032229407224804163
step: 220, loss: 0.007276154588907957
step: 230, loss: 0.01500729750841856
step: 240, loss: 0.019522592425346375
step: 250, loss: 0.01878364197909832
```

```
step: 260, loss: 0.006786208134144545
step: 270, loss: 0.0002125797764165327
step: 280, loss: 0.004267312120646238
step: 290, loss: 0.09476140886545181
step: 300, loss: 0.0021209127735346556
step: 310, loss: 0.00023385082022286952
step: 320, loss: 0.029833802953362465
step: 330, loss: 0.002774506341665983
step: 340, loss: 0.010658608749508858
step: 350, loss: 0.006102377083152533
step: 360, loss: 0.004166932310909033
step: 370, loss: 0.00023761013289913535
step: 380, loss: 0.0023445766419172287
step: 390, loss: 0.0018215967575088143
step: 400, loss: 0.037627480924129486
step: 410, loss: 0.0040895682759583
step: 420, loss: 0.0005273489514365792
step: 430, loss: 0.002296384423971176
step: 440, loss: 0.0008560302667319775
step: 450, loss: 0.0030368922743946314
step: 460, loss: 0.00163132487796247
step: 470, loss: 0.034151606261730194
step: 480, loss: 0.006359501276165247
step: 490, loss: 0.00024916810798458755
step: 500, loss: 0.0002724384830798954
step: 510, loss: 0.0006992605049163103
step: 520, loss: 0.25812891125679016
step: 530, loss: 0.008803033269941807
/usr/local/lib/python3.7/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=0.8995305164319248, f1=0.8838311996206734,
best_f1=0.9104477611940298
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0010449629044160247
step: 10, loss: 0.0008594889659434557
step: 20, loss: 0.0015271547017619014
step: 30, loss: 0.0007915567839518189
step: 40, loss: 0.02400072291493416
step: 50, loss: 0.0008694994612596929
step: 60, loss: 0.002102307043969631
step: 70, loss: 0.005107145290821791
step: 80, loss: 0.0013852114789187908
step: 90, loss: 8.235334098571911e-05
step: 100, loss: 0.014856306836009026
step: 110, loss: 0.0004340430023148656
```

```
step: 120, loss: 0.00042558545828796923
step: 130, loss: 0.0019454753492027521
step: 140, loss: 0.00010209379252046347
step: 150, loss: 0.0028622914105653763
step: 160, loss: 0.0053033907897770405
step: 170, loss: 0.00026179637643508613
step: 180, loss: 0.0005481241387315094
step: 190, loss: 0.0054372940212488174
step: 200, loss: 0.0002386776322964579
step: 210, loss: 0.07276497036218643
step: 220, loss: 0.0021834198851138353
step: 230, loss: 0.003547589760273695
step: 240, loss: 0.07326788455247879
step: 250, loss: 0.008416907861828804
step: 260, loss: 0.002834924729540944
step: 270, loss: 0.00011732481652870774
step: 280, loss: 0.028157051652669907
step: 290, loss: 0.004547297954559326
step: 300, loss: 0.00041533485637046397
step: 310, loss: 0.00109567039180547
step: 320, loss: 0.0004044041852466762
step: 330, loss: 0.003341520670801401
step: 340, loss: 0.002875254489481449
step: 350, loss: 0.0008203869801945984
step: 360, loss: 0.000877460406627506
step: 370, loss: 0.0001723577152006328
step: 380, loss: 0.00011180635920027271
step: 390, loss: 0.0005369606078602374
step: 400, loss: 0.011992468498647213
step: 410, loss: 0.00012193244037916884
step: 420, loss: 0.0004111335729248822
step: 430, loss: 0.00017416418995708227
step: 440, loss: 0.0025300111155956984
step: 450, loss: 0.0013613927876576781
step: 460, loss: 0.0009743022383190691
step: 470, loss: 0.00026293261907994747
step: 480, loss: 0.00021200325863901526
step: 490, loss: 0.00012393620272632688
step: 500, loss: 0.0002930932096205652
step: 510, loss: 0.0003397312539163977
step: 520, loss: 0.044243164360523224
step: 530, loss: 0.0022585734259337187
/usr/local/lib/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=0.9099264705882353, f1=0.9010082493125573,
best_f1=0.9010082493125573
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00012710157898254693
step: 10, loss: 0.001811484806239605
step: 20, loss: 0.00025207403814420104
step: 30, loss: 0.04580257833003998
step: 40, loss: 0.0017608747584745288
step: 50, loss: 0.00016661181871313602
step: 60, loss: 0.0013245880836620927
step: 70, loss: 0.09608511626720428
step: 80, loss: 5.836014315718785e-05
step: 90, loss: 0.00696758134290576
step: 100, loss: 0.00011815282050520182
step: 110, loss: 0.0002632666437420994
step: 120, loss: 6.68893480906263e-05
step: 130, loss: 0.005571961402893066
step: 140, loss: 0.00019173239707015455
step: 150, loss: 0.00010960292274830863
step: 160, loss: 9.629800479160622e-05
step: 170, loss: 8.799450733931735e-05
step: 180, loss: 0.009281539358198643
step: 190, loss: 0.00016796053387224674
step: 200, loss: 0.00016385194612666965
step: 210, loss: 0.00019912865536753088
step: 220, loss: 0.00015778363740537316
step: 230, loss: 0.0053124018013477325
step: 240, loss: 0.0004940003855153918
step: 250, loss: 0.005892560351639986
step: 260, loss: 0.00012653916201088578
step: 270, loss: 0.0013661248376592994
step: 280, loss: 7.181961700553074e-05
step: 290, loss: 0.0001340630988124758
step: 300, loss: 0.0001812126865843311
step: 310, loss: 0.0005185390473343432
step: 320, loss: 0.009207877330482006
step: 330, loss: 7.320723671000451e-05
step: 340, loss: 0.22737742960453033
step: 350, loss: 0.00010987262066919357
step: 360, loss: 0.06278065592050552
step: 370, loss: 0.0013377299765124917
step: 380, loss: 0.0004901491920463741
step: 390, loss: 0.00016655886429362
step: 400, loss: 0.0012883845483884215
step: 410, loss: 0.030197497457265854
```

```
step: 420, loss: 0.0005196749698370695
step: 430, loss: 0.0007776746060699224
step: 440, loss: 0.05967879667878151
step: 450, loss: 7.489875133614987e-05
step: 460, loss: 0.0002458388917148113
step: 470, loss: 8.382445957977325e-05
step: 480, loss: 0.00016517947369720787
step: 490, loss: 7.674522930756211e-05
step: 500, loss: 0.0001734695688355714
step: 510, loss: 0.0010884073562920094
step: 520, loss: 8.742558566154912e-05
step: 530, loss: 0.00022627795988228172
/usr/local/lib/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=0.9158530915853093, f1=0.9069767441860467,
best_f1=0.9069767441860467
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.031125258654356003
step: 10, loss: 5.099217014503665e-05
step: 20, loss: 7.328562787733972e-05
step: 30, loss: 5.222554682404734e-05
step: 40, loss: 0.000323205633321777
step: 50, loss: 0.00023811060236766934
step: 60, loss: 0.00016147579299286008
step: 70, loss: 0.00046902187750674784
step: 80, loss: 0.001881958800368011
step: 90, loss: 0.0004419989127200097
step: 100, loss: 5.053061613580212e-05
step: 110, loss: 0.0007081162184476852
step: 120, loss: 4.041002466692589e-05
step: 130, loss: 0.0001469740818720311
step: 140, loss: 0.00010058742191176862
step: 150, loss: 0.00033867047750391066
step: 160, loss: 0.0002014257333939895
step: 170, loss: 0.00023553776554763317
step: 180, loss: 0.01798698864877224
step: 190, loss: 0.0001193598291138187
step: 200, loss: 0.02268211543560028
step: 210, loss: 0.006876681931316853
step: 220, loss: 0.0006023194873705506
step: 230, loss: 0.000272151519311592
```

```
step: 240, loss: 0.0003503556945361197
step: 250, loss: 0.00041474390309304
step: 260, loss: 0.0007429607794620097
step: 270, loss: 0.001692013582214713
step: 280, loss: 0.00010825843492057174
step: 290, loss: 5.856111602042802e-05
step: 300, loss: 0.0005933787324465811
step: 310, loss: 0.025121593847870827
step: 320, loss: 0.000229112891247496
step: 330, loss: 0.012972884811460972
step: 340, loss: 0.009888997301459312
step: 350, loss: 0.0004994641640223563
step: 360, loss: 9.330619650427252e-05
step: 370, loss: 0.0020957186352461576
step: 380, loss: 0.00029809516854584217
step: 390, loss: 0.00020150220370851457
step: 400, loss: 0.0003562269266694784
step: 410, loss: 0.0010880044428631663
step: 420, loss: 0.007383884862065315
step: 430, loss: 0.00012508462532423437
step: 440, loss: 7.621372060384601e-05
step: 450, loss: 0.004394313786178827
step: 460, loss: 0.00016555498586967587
step: 470, loss: 0.00010418774036224931
step: 480, loss: 5.058643000666052e-05
step: 490, loss: 0.004938443657010794
step: 500, loss: 0.0008587267948314548
step: 510, loss: 0.0005562136648222804
step: 520, loss: 0.0009345613652840257
step: 530, loss: 0.0019049689872190356
/usr/local/lib/python3.7/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=0.904476234425473, f1=0.8968070337806572,
best_f1=0.9069767441860467
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001654300489462912
step: 10, loss: 0.0020877299830317497
step: 20, loss: 4.7750720113981515e-05
step: 30, loss: 0.019613126292824745
step: 40, loss: 0.0011956198140978813
step: 50, loss: 0.09767451882362366
step: 60, loss: 0.00029963150154799223
step: 70, loss: 0.000275350728770718
step: 80, loss: 0.000358305755071342
step: 90, loss: 0.00025505467783659697
```

```
step: 100, loss: 0.0005356220644898713
step: 110, loss: 0.00024373221094720066
step: 120, loss: 0.0003491388924885541
step: 130, loss: 0.013229712843894958
step: 140, loss: 4.205533332424238e-05
step: 150, loss: 0.002645474160090089
step: 160, loss: 0.0006867969059385359
step: 170, loss: 0.00018871112843044102
step: 180, loss: 0.0001726567861624062
step: 190, loss: 0.0009360762778669596
step: 200, loss: 0.000986530794762075
step: 210, loss: 0.00021528905199375004
step: 220, loss: 7.62420822866261e-05
step: 230, loss: 0.0003438446728978306
step: 240, loss: 5.6092889280989766e-05
step: 250, loss: 0.001011133543215692
step: 260, loss: 9.591527486918494e-05
step: 270, loss: 0.002695602597668767
step: 280, loss: 0.0002744044177234173
step: 290, loss: 0.0010979297803714871
step: 300, loss: 0.0022948586847633123
step: 310, loss: 0.016274062916636467
step: 320, loss: 0.0008773254230618477
step: 330, loss: 0.00041663035517558455
step: 340, loss: 0.00030127938953228295
step: 350, loss: 0.0001914358581416309
step: 360, loss: 0.0001377103035338223
step: 370, loss: 4.1369425161974505e-05
step: 380, loss: 0.00034361210418865085
step: 390, loss: 0.0030731225851923227
step: 400, loss: 4.901912689092569e-05
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 410, loss: 0.08131949603557587
step: 420, loss: 0.025283444672822952
step: 430, loss: 0.0031803783494979143
step: 440, loss: 6.969686364755034e-05
step: 450, loss: 4.9273829063167796e-05
step: 460, loss: 0.001445887959562242
step: 470, loss: 0.0022307501640170813
step: 480, loss: 5.057920134277083e-05
step: 490, loss: 0.0004650955379474908
step: 500, loss: 5.539812264032662e-05
step: 510, loss: 2.630011113069486e-05
step: 520, loss: 0.00023847923148423433
step: 530, loss: 2.9905300834798254e-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=0.9074327405380758, f1=0.9032553874369554,
best_f1=0.9069767441860467
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.091176379006356e-05
step: 10, loss: 3.366453529451974e-05
step: 20, loss: 0.0008462695986963809
step: 30, loss: 0.00025200253003276885
step: 40, loss: 0.0005262693157419562
step: 50, loss: 3.0107115890132263e-05
step: 60, loss: 3.827879118034616e-05
step: 70, loss: 4.6601078793173656e-05
step: 80, loss: 0.0005695407744497061
step: 90, loss: 3.5482371458783746e-05
step: 100, loss: 0.00011210536467842758
step: 110, loss: 4.1090359445661306e-05
step: 120, loss: 0.02698628418147564
step: 130, loss: 0.00011947994789807126
step: 140, loss: 5.200216037337668e-05
step: 150, loss: 0.00038717302959412336
step: 160, loss: 0.0014169042697176337
step: 170, loss: 5.9110665461048484e-05
step: 180, loss: 3.4148335544159636e-05
step: 190, loss: 4.252750659361482e-05
step: 200, loss: 5.074121509096585e-05
step: 210, loss: 6.444691098295152e-05
step: 220, loss: 3.3299438655376434e-05
step: 230, loss: 0.0035138323437422514
step: 240, loss: 9.688379941508174e-05
step: 250, loss: 0.00010238060349365696
step: 260, loss: 0.00028523773653432727
step: 270, loss: 5.356501787900925e-05
step: 280, loss: 0.013149772770702839
step: 290, loss: 0.0011182422749698162
step: 300, loss: 0.00015541893662884831
step: 310, loss: 8.448466542176902e-05
step: 320, loss: 0.0005379998474381864
step: 330, loss: 3.159328844049014e-05
step: 340, loss: 7.068787090247497e-05
step: 350, loss: 0.0005230975802987814
step: 360, loss: 0.0004228855250403285
step: 370, loss: 9.188630065182224e-05
step: 380, loss: 4.1881692595779896e-05
step: 390, loss: 4.305052061681636e-05
step: 400, loss: 0.001756215700879693
step: 410, loss: 0.04588715359568596
step: 420, loss: 0.012607629410922527
```

```
step: 430, loss: 0.008855445310473442
step: 440, loss: 0.0009375223889946938
step: 450, loss: 0.0005675139254890382
step: 460, loss: 0.0019415965070948005
step: 470, loss: 0.0003656769695226103
step: 480, loss: 4.466063910513185e-05
step: 490, loss: 7.784023182466626e-05
step: 500, loss: 0.006576853338629007
step: 510, loss: 4.883242945652455e-05
step: 520, loss: 0.005180027801543474
step: 530, loss: 0.0014334887964650989
/usr/local/lib/python3.7/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=0.9016166281755196, f1=0.8931050439611292,
best_f1=0.9069767441860467
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.605150413932279e-05
step: 10, loss: 3.612908039940521e-05
step: 20, loss: 9.746100113261491e-05
step: 30, loss: 0.0021923540625721216
step: 40, loss: 4.9667327402858064e-05
step: 50, loss: 3.0241146305343136e-05
step: 60, loss: 3.896986527252011e-05
step: 70, loss: 3.39209582307376e-05
step: 80, loss: 0.0045065609738230705
step: 90, loss: 0.00022525039094034582
step: 100, loss: 0.0006386112654581666
step: 110, loss: 0.012835056520998478
step: 120, loss: 0.0019952021539211273
step: 130, loss: 0.004622084088623524
step: 140, loss: 0.0011830731527879834
step: 150, loss: 2.9879747671657242e-05
step: 160, loss: 4.200092735118233e-05
step: 170, loss: 6.403511360986158e-05
step: 180, loss: 0.0037274810019880533
step: 190, loss: 3.713112892000936e-05
step: 200, loss: 0.0012100718449801207
step: 210, loss: 0.00025412882678210735
step: 220, loss: 2.3502374460804276e-05
step: 230, loss: 7.677119720028713e-05
step: 240, loss: 3.311291220597923e-05
step: 250, loss: 0.0009444250026717782
step: 260, loss: 0.00012453034287318587
step: 270, loss: 3.200252467649989e-05
step: 280, loss: 1.991513818211388e-05
```

```
step: 290, loss: 3.776461380766705e-05
step: 300, loss: 3.578010364435613e-05
step: 310, loss: 4.10159227612894e-05
step: 320, loss: 3.083683259319514e-05
step: 330, loss: 4.296115366742015e-05
step: 340, loss: 2.9142203857190907e-05
step: 350, loss: 3.627121259341948e-05
step: 360, loss: 2.1561605535680428e-05
step: 370, loss: 2.8977185138501227e-05
step: 380, loss: 0.000666355132125318
step: 390, loss: 4.466906102607027e-05
step: 400, loss: 2.055214463325683e-05
step: 410, loss: 0.00011210676893824711
step: 420, loss: 2.179258081014268e-05
step: 430, loss: 0.021266089752316475
step: 440, loss: 0.021105865016579628
step: 450, loss: 2.79838168353308e-05
step: 460, loss: 4.19963471358642e-05
step: 470, loss: 0.003944851458072662
step: 480, loss: 0.00137917953543365
step: 490, loss: 0.0007383065531030297
step: 500, loss: 0.00013576331548392773
step: 510, loss: 2.7685740860761143e-05
step: 520, loss: 0.0046376436948776245
step: 530, loss: 4.1287978092441335e-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=0.9041474654377881, f1=0.8934010152284264,
best_f1=0.9069767441860467
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.867347135790624e-05
step: 10, loss: 0.0002642383915372193
step: 20, loss: 9.620797936804593e-05
step: 30, loss: 0.0011449161684140563
step: 40, loss: 5.538840559893288e-05
step: 50, loss: 3.3723441447364166e-05
step: 60, loss: 4.47359707322903e-05
step: 70, loss: 2.1885778551222757e-05
step: 80, loss: 5.0568985898280516e-05
step: 90, loss: 0.002562990179285407
step: 100, loss: 6.420002318918705e-05
step: 110, loss: 8.522329881088808e-05
step: 120, loss: 0.0001976750500034541
step: 130, loss: 0.000411076529417187
step: 140, loss: 0.013179624453186989
```

```
step: 150, loss: 0.000386564526706934
step: 160, loss: 0.0008030469762161374
step: 170, loss: 3.6775105400010943e-05
step: 180, loss: 3.814530646195635e-05
step: 190, loss: 0.0084312055259943
step: 200, loss: 5.866992796654813e-05
step: 210, loss: 0.0024407743476331234
step: 220, loss: 0.0012374832294881344
step: 230, loss: 0.000518572167493403
step: 240, loss: 3.49384754372295e-05
step: 250, loss: 6.222239608177915e-05
step: 260, loss: 2.603892789920792e-05
step: 270, loss: 0.004786367062479258
step: 280, loss: 0.002147115534171462
step: 290, loss: 0.00011210680531803519
step: 300, loss: 0.00026087561855092645
step: 310, loss: 3.5295586712891236e-05
step: 320, loss: 5.711897392757237e-05
step: 330, loss: 0.00020519208919722587
step: 340, loss: 0.0002542723377700895
step: 350, loss: 0.00011697415902744979
step: 360, loss: 0.00015846570022404194
step: 370, loss: 0.0001857835886767134
step: 380, loss: 4.407556480146013e-05
step: 390, loss: 3.223751627956517e-05
step: 400, loss: 3.0162802431732416e-05
step: 410, loss: 4.318806895753369e-05
step: 420, loss: 4.84569973195903e-05
step: 430, loss: 2.798011337290518e-05
step: 440, loss: 2.1099778678035364e-05
step: 450, loss: 2.9745629944955e-05
step: 460, loss: 2.7939129722653888e-05
step: 470, loss: 5.590490036411211e-05
step: 480, loss: 0.0011744922958314419
step: 490, loss: 3.8958100049057975e-05
step: 500, loss: 0.0001590478786965832
step: 510, loss: 0.00019903035718016326
step: 520, loss: 5.844482438988052e-05
step: 530, loss: 0.0003778412065003067
/usr/local/lib/python3.7/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=0.9079189686924493, f1=0.8976014760147601,
best_f1=0.9069767441860467
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003732895420398563
```

```
step: 10, loss: 7.040593482088298e-05
step: 20, loss: 4.4415679440135136e-05
step: 30, loss: 0.0001274504465982318
step: 40, loss: 2.2113010345492512e-05
step: 50, loss: 5.436444189399481e-05
step: 60, loss: 2.8605647457879968e-05
step: 70, loss: 5.6764623877825215e-05
step: 80, loss: 3.646118057076819e-05
step: 90, loss: 4.877339961240068e-05
step: 100, loss: 2.850152122846339e-05
step: 110, loss: 4.747490675072186e-05
step: 120, loss: 0.0016623721458017826
step: 130, loss: 0.00027713290182873607
step: 140, loss: 2.2854314011055976e-05
step: 150, loss: 5.511810741154477e-05
step: 160, loss: 0.0001222019927809015
step: 170, loss: 3.357484092703089e-05
step: 180, loss: 0.00014840011135675013
step: 190, loss: 5.69140502193477e-05
step: 200, loss: 2.5290555640822276e-05
step: 210, loss: 0.0025503505021333694
step: 220, loss: 0.00034243383561261
step: 230, loss: 0.0005987408221699297
step: 240, loss: 0.0001609615283086896
step: 250, loss: 0.0059795803390443325
step: 260, loss: 3.868846033583395e-05
step: 270, loss: 9.815231169341132e-05
step: 280, loss: 4.259631532477215e-05
step: 290, loss: 0.0033774932380765676
step: 300, loss: 0.00011097986134700477
step: 310, loss: 0.0010117068886756897
step: 320, loss: 5.384182441048324e-05
step: 330, loss: 0.001377042499370873
step: 340, loss: 0.009545385837554932
step: 350, loss: 0.0003916913701687008
step: 360, loss: 7.666352030355483e-05
step: 370, loss: 2.0302586563047953e-05
step: 380, loss: 2.17217984754825e-05
step: 390, loss: 2.2034775611246005e-05
step: 400, loss: 2.314112862222828e-05
step: 410, loss: 0.00023148178297560662
step: 420, loss: 3.3984466426773e-05
step: 430, loss: 3.065780038014054e-05
step: 440, loss: 7.525619730586186e-05
step: 450, loss: 4.537457425612956e-05
step: 460, loss: 8.093791257124394e-05
step: 470, loss: 3.870565342367627e-05
step: 480, loss: 5.521561615751125e-05
```

```
step: 490, loss: 3.567506428225897e-05
step: 500, loss: 2.6612977308104746e-05
step: 510, loss: 3.697119245771319e-05
step: 520, loss: 2.2463173081632704e-05
step: 530, loss: 4.254799205227755e-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=0.9100968188105117, f1=0.9000925069380203,
best_f1=0.9069767441860467
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.")
5742it [00:23, 245.70it/s]
load_f1 = 0.9054494643688868
real_f1 = 0.9017233348858873
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, 254.54it/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: 0.8338878154754639
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.06497710198163986
step: 20, loss: 0.38273608684539795
step: 30, loss: 0.37137657403945923
step: 40, loss: 0.5108407735824585
step: 50, loss: 0.3030611276626587
step: 60, loss: 0.37115228176116943
step: 70, loss: 0.2813487648963928
step: 80, loss: 0.2790031135082245
step: 90, loss: 0.4297909140586853
step: 100, loss: 0.20846620202064514
step: 110, loss: 0.3076065182685852
step: 120, loss: 0.26456400752067566
step: 130, loss: 0.21347135305404663
step: 140, loss: 0.3191094398498535
step: 150, loss: 0.3268556296825409
step: 160, loss: 0.26260805130004883
step: 170, loss: 0.18959330022335052
step: 180, loss: 0.19118285179138184
step: 190, loss: 0.2612443268299103
step: 200, loss: 0.15704479813575745
step: 210, loss: 0.47511595487594604
/usr/local/lib/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=0.3824451410658307, f1=0.38880248833592534,
best_f1=0.38880248833592534
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09353360533714294
step: 10, loss: 0.1322459727525711
step: 20, loss: 0.30116644501686096
step: 30, loss: 0.24740764498710632
step: 40, loss: 0.12598305940628052
step: 50, loss: 0.1700034886598587
step: 60, loss: 0.11516140401363373
step: 70, loss: 0.24956417083740234
step: 80, loss: 0.28106510639190674
step: 90, loss: 0.14641375839710236
step: 100, loss: 0.15647655725479126
step: 110, loss: 0.06928884238004684
step: 120, loss: 0.24786721169948578
step: 130, loss: 0.2297855019569397
step: 140, loss: 0.28161945939064026
step: 150, loss: 0.32324904203414917
step: 160, loss: 0.17388340830802917
step: 170, loss: 0.23023951053619385
step: 180, loss: 0.44579562544822693
step: 190, loss: 0.20656836032867432
step: 200, loss: 0.22039370238780975
step: 210, loss: 0.284611314535141
/usr/local/lib/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=0.49900990099009895, f1=0.49218750000000006,
best_f1=0.49218750000000006
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.29266121983528137
step: 10, loss: 0.24167804419994354
step: 20, loss: 0.32012367248535156
step: 30, loss: 0.08076155930757523
```

```
step: 40, loss: 0.1558769792318344
step: 50, loss: 0.17722167074680328
step: 60, loss: 0.2881155014038086
step: 70, loss: 0.26495757699012756
step: 80, loss: 0.15170399844646454
step: 90, loss: 0.17096436023712158
step: 100, loss: 0.14330120384693146
step: 110, loss: 0.2530754804611206
step: 120, loss: 0.2627081871032715
step: 130, loss: 0.23430706560611725
step: 140, loss: 0.2839588522911072
step: 150, loss: 0.1559518426656723
step: 160, loss: 0.09332317858934402
step: 170, loss: 0.15985563397407532
step: 180, loss: 0.08176379650831223
step: 190, loss: 0.2081272155046463
step: 200, loss: 0.19137969613075256
step: 210, loss: 0.247709721326828
/usr/local/lib/python3.7/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=0.46399999999999997, f1=0.48992248062015503,
best_f1=0.49218750000000006
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.19393259286880493
step: 10, loss: 0.07184434682130814
step: 20, loss: 0.14077900350093842
step: 30, loss: 0.05794326588511467
step: 40, loss: 0.0845976322889328
step: 50, loss: 0.1028946116566658
step: 60, loss: 0.08110840618610382
step: 70, loss: 0.18115536868572235
step: 80, loss: 0.11933949589729309
step: 90, loss: 0.11839695274829865
step: 100, loss: 0.2508111000061035
step: 110, loss: 0.17835280299186707
step: 120, loss: 0.07458696514368057
step: 130, loss: 0.12932446599006653
step: 140, loss: 0.20528781414031982
step: 150, loss: 0.3058883845806122
step: 160, loss: 0.12010950595140457
step: 170, loss: 0.07084677368402481
step: 180, loss: 0.16082356870174408
step: 190, loss: 0.1550751030445099
step: 200, loss: 0.30840063095092773
step: 210, loss: 0.028456291183829308
```

```
/usr/local/lib/python3.7/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=0.4892086330935252, f1=0.5045372050816697,
best_f1=0.49218750000000006
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0319003164768219
step: 10, loss: 0.024810025468468666
step: 20, loss: 0.03419032692909241
step: 30, loss: 0.052991751581430435
step: 40, loss: 0.11496379226446152
step: 50, loss: 0.10563930869102478
step: 60, loss: 0.012383995577692986
step: 70, loss: 0.05897487699985504
step: 80, loss: 0.03536815196275711
step: 90, loss: 0.032101113349199295
step: 100, loss: 0.051147595047950745
step: 110, loss: 0.011913386173546314
step: 120, loss: 0.03212032467126846
step: 130, loss: 0.09798366576433182
step: 140, loss: 0.08156851679086685
step: 150, loss: 0.06785580515861511
step: 160, loss: 0.07603263854980469
step: 170, loss: 0.16761212050914764
step: 180, loss: 0.17829345166683197
step: 190, loss: 0.035667866468429565
step: 200, loss: 0.10190059244632721
step: 210, loss: 0.10239781439304352
/usr/local/lib/python3.7/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=0.4794007490636704, f1=0.5057471264367815,
best_f1=0.49218750000000006
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0782371535897255
step: 10, loss: 0.010946298018097878
step: 20, loss: 0.02598969079554081
step: 30, loss: 0.09672939777374268
step: 40, loss: 0.03216692432761192
step: 50, loss: 0.08070214837789536
step: 60, loss: 0.15275838971138
step: 70, loss: 0.0017126903403550386
step: 80, loss: 0.0703752413392067
step: 90, loss: 0.08091215789318085
```

```
step: 100, loss: 0.013807358220219612
step: 110, loss: 0.018015142530202866
step: 120, loss: 0.047368474304676056
step: 130, loss: 0.01117868721485138
step: 140, loss: 0.10730523616075516
step: 150, loss: 0.12075506895780563
step: 160, loss: 0.13373751938343048
step: 170, loss: 0.08139638602733612
step: 180, loss: 0.012852126732468605
step: 190, loss: 0.07666759938001633
step: 200, loss: 0.01204307284206152
step: 210, loss: 0.04715175926685333
/usr/local/lib/python3.7/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=0.4387755102040817, f1=0.47750865051903113,
best_f1=0.49218750000000006
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.017000848427414894
step: 10, loss: 0.027916939929127693
step: 20, loss: 0.07492143660783768
step: 30, loss: 0.005855211056768894
step: 40, loss: 0.009085793979465961
step: 50, loss: 0.040298763662576675
step: 60, loss: 0.20512782037258148
step: 70, loss: 0.007256664801388979
step: 80, loss: 0.017021574079990387
step: 90, loss: 0.0072993142530322075
step: 100, loss: 0.04362857714295387
step: 110, loss: 0.09218839555978775
step: 120, loss: 0.007197810336947441
step: 130, loss: 0.0029788960237056017
step: 140, loss: 0.0008066437439993024
step: 150, loss: 0.02458552084863186
step: 160, loss: 0.04135756567120552
step: 170, loss: 0.03155176341533661
step: 180, loss: 0.004753591027110815
step: 190, loss: 0.057393159717321396
step: 200, loss: 0.04294731095433235
step: 210, loss: 0.040976013988256454
/usr/local/lib/python3.7/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=0.43963963963963965, f1=0.48888888888888893,
best_f1=0.49218750000000006
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

```
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012109278701245785
step: 10, loss: 0.011751762591302395
step: 20, loss: 0.009209365583956242
step: 30, loss: 0.0014716251753270626
step: 40, loss: 0.030257536098361015
step: 50, loss: 0.02277670055627823
step: 60, loss: 0.23048779368400574
step: 70, loss: 0.009943751618266106
step: 80, loss: 0.04202204570174217
step: 90, loss: 0.046513818204402924
step: 100, loss: 0.012660510838031769
step: 110, loss: 0.008888518437743187
step: 120, loss: 0.006271565333008766
step: 130, loss: 0.05151132494211197
step: 140, loss: 0.007443216163665056
step: 150, loss: 0.06714180111885071
step: 160, loss: 0.11897037923336029
step: 170, loss: 0.0009770413162186742
step: 180, loss: 0.028074558824300766
step: 190, loss: 0.02914891391992569
step: 200, loss: 0.10696481913328171
step: 210, loss: 0.07837813347578049
/usr/local/lib/python3.7/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=0.4572564612326043, f1=0.49896907216494846,
best_f1=0.49218750000000006
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.017119556665420532
step: 10, loss: 0.003756521502509713
step: 20, loss: 0.04297392815351486
step: 30, loss: 0.030472377315163612
step: 40, loss: 0.07798801362514496
step: 50, loss: 0.007486653048545122
step: 60, loss: 0.0010313253151252866
step: 70, loss: 0.008587422780692577
step: 80, loss: 0.0012304057599976659
step: 90, loss: 0.10415980219841003
step: 100, loss: 0.0032600234262645245
step: 110, loss: 0.0007691835053265095
step: 120, loss: 0.001676974119618535
step: 130, loss: 0.003896548645570874
step: 140, loss: 0.0077142538502812386
step: 150, loss: 0.0617019459605217
```

```
step: 160, loss: 0.00028995683533139527
step: 170, loss: 0.0020947307348251343
step: 180, loss: 0.015116715803742409
step: 190, loss: 0.05891421437263489
step: 200, loss: 0.01300676167011261
step: 210, loss: 0.05510669946670532
/usr/local/lib/python3.7/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=0.45783132530120485, f1=0.4696356275303644,
best_f1=0.49218750000000006
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002767236903309822
step: 10, loss: 0.004063886124640703
step: 20, loss: 0.017100142315030098
step: 30, loss: 0.003199910745024681
step: 40, loss: 0.04819553345441818
step: 50, loss: 0.006192728877067566
step: 60, loss: 0.0005233351839706302
step: 70, loss: 0.0016189905581995845
step: 80, loss: 0.000508347584400326
step: 90, loss: 0.006022964604198933
step: 100, loss: 0.0008187335333786905
step: 110, loss: 0.0028321873396635056
step: 120, loss: 0.0008070819894783199
step: 130, loss: 0.0008456273353658617
step: 140, loss: 0.0002445665595587343
step: 150, loss: 0.15087342262268066
step: 160, loss: 0.017695432528853416
step: 170, loss: 0.002236902480944991
step: 180, loss: 0.0012027439661324024
step: 190, loss: 0.019211288541555405
step: 200, loss: 0.004555991850793362
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 210, loss: 0.004717202391475439
/usr/local/lib/python3.7/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=0.4463894967177243, f1=0.43859649122807015,
best_f1=0.49218750000000006
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02496843785047531
step: 10, loss: 0.004682432860136032
step: 20, loss: 0.06344481557607651
```

```
step: 30, loss: 0.03850361332297325
step: 40, loss: 0.005555060226470232
step: 50, loss: 0.0006960831233300269
step: 60, loss: 0.00028670954634435475
step: 70, loss: 0.002888414543122053
step: 80, loss: 0.09255292266607285
step: 90, loss: 0.0010166139109060168
step: 100, loss: 0.0009217164479196072
step: 110, loss: 0.003042178228497505
step: 120, loss: 0.002161058597266674
step: 130, loss: 0.03051399067044258
step: 140, loss: 0.048966072499752045
step: 150, loss: 0.005869658198207617
step: 160, loss: 0.0014246719656512141
step: 170, loss: 0.1310221552848816
step: 180, loss: 0.0035930778831243515
step: 190, loss: 0.0046998606994748116
step: 200, loss: 0.00021208437101449817
step: 210, loss: 0.0005235841381363571
/usr/local/lib/python3.7/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=0.4375, f1=0.4403292181069959, best_f1=0.49218750000000006
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00028210721211507916
step: 10, loss: 0.0012637312756851315
step: 20, loss: 0.011408167891204357
step: 30, loss: 0.02085551992058754
step: 40, loss: 0.05751106142997742
step: 50, loss: 0.0010013204300776124
step: 60, loss: 0.009316311217844486
step: 70, loss: 0.0012404231820255518
step: 80, loss: 0.0006599728367291391
step: 90, loss: 0.0004346598288975656
step: 100, loss: 0.00018284816178493202
step: 110, loss: 0.0011865299893543124
step: 120, loss: 0.0036112440284341574
step: 130, loss: 0.001374050509184599
step: 140, loss: 0.0012604845687747002
step: 150, loss: 0.00017267794464714825
step: 160, loss: 0.0008969481568783522
step: 170, loss: 0.00033243693178519607
step: 180, loss: 0.005259924568235874
step: 190, loss: 0.0006858069682493806
step: 200, loss: 0.0011195304105058312
step: 210, loss: 0.002078506164252758
```

```
/usr/local/lib/python3.7/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=0.3981264637002342, f1=0.4232558139534883,
best_f1=0.49218750000000006
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.839755562599748e-05
step: 10, loss: 0.000406853825552389
step: 20, loss: 0.00027645481168292463
step: 30, loss: 0.0005868393345735967
step: 40, loss: 0.0007998545770533383
step: 50, loss: 0.003176879370585084
step: 60, loss: 0.00015675624308642
step: 70, loss: 0.048018135130405426
step: 80, loss: 0.00021798143279738724
step: 90, loss: 0.03358355909585953
step: 100, loss: 0.052028853446245193
step: 110, loss: 0.000436162983532995
step: 120, loss: 0.0006057150894775987
step: 130, loss: 0.000510345445945859
step: 140, loss: 0.13916811347007751
step: 150, loss: 0.09164286404848099
step: 160, loss: 0.002156950067728758
step: 170, loss: 0.0002371268637944013
step: 180, loss: 0.0004598866798914969
step: 190, loss: 0.00032123998971655965
step: 200, loss: 0.00039617778384126723
step: 210, loss: 0.0016223834827542305
/usr/local/lib/python3.7/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=0.43933054393305443, f1=0.4458333333333333,
best_f1=0.49218750000000006
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004745836369693279
step: 10, loss: 0.0009157445165328681
step: 20, loss: 0.002135121962055564
step: 30, loss: 0.006142637692391872
step: 40, loss: 0.0016895137960091233
step: 50, loss: 0.0009162202477455139
step: 60, loss: 0.004152150359004736
step: 70, loss: 0.0011472722981125116
step: 80, loss: 0.0019432558910921216
step: 90, loss: 0.0008832734310999513
```

```
step: 100, loss: 0.00024872971698641777
step: 110, loss: 0.0005269584944471717
step: 120, loss: 0.0006313372869044542
step: 130, loss: 0.0005839674849994481
step: 140, loss: 0.0014074043137952685
step: 150, loss: 0.0006458775606006384
step: 160, loss: 0.01908164471387863
step: 170, loss: 0.01376801822334528
step: 180, loss: 0.0003160179185215384
step: 190, loss: 0.00011537795217009261
step: 200, loss: 0.0002078519610222429
step: 210, loss: 0.0004948569112457335
/usr/local/lib/python3.7/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=0.438177874186551, f1=0.44347826086956516,
best_f1=0.49218750000000006
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005284194485284388
step: 10, loss: 0.00047788457595574324
step: 20, loss: 0.013378809206187725
step: 30, loss: 0.0005869587766937912
step: 40, loss: 0.00022351212101057172
step: 50, loss: 0.00010028070391854271
step: 60, loss: 0.0002727008832152933
step: 70, loss: 0.0004395965952426195
step: 80, loss: 0.00026717365835793316
step: 90, loss: 0.00041518916259519756
step: 100, loss: 0.00011150990758324042
step: 110, loss: 0.00013876281445845962
step: 120, loss: 0.0006130276597104967
step: 130, loss: 0.016525575891137123
step: 140, loss: 0.000304974993923679
step: 150, loss: 0.004055396653711796
step: 160, loss: 0.0036864825524389744
step: 170, loss: 0.1351076364517212
step: 180, loss: 0.024800432845950127
step: 190, loss: 0.00043292134068906307
step: 200, loss: 0.006173199974000454
step: 210, loss: 0.0071951416321098804
/usr/local/lib/python3.7/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=0.4298245614035088, f1=0.4410480349344978,
best_f1=0.49218750000000006
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.")
2293it [00:06, 330.55it/s]
load_f1 = 0.4724061810154525
real_f1 = 0.4794816414686825
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, 256.75it/s]
```

3.1.5 Walmart-Amazon - Running the matcher

```python
[ ]: # 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 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: 0.8501303791999817
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.16457441449165344
step: 20, loss: 0.15011945366859436
step: 30, loss: 0.5152413249015808
step: 40, loss: 0.26419398188591003
step: 50, loss: 0.3114648163318634
step: 60, loss: 0.36679893732070923
step: 70, loss: 0.17469274997711182
step: 80, loss: 0.5372204780578613
step: 90, loss: 0.25185471773147583
step: 100, loss: 0.22086888551712036
step: 110, loss: 0.23496046662330627
step: 120, loss: 0.42458176612854004
step: 130, loss: 0.34671908617019653
step: 140, loss: 0.3412063717842102
step: 150, loss: 0.2744591236114502
step: 160, loss: 0.20913326740264893
step: 170, loss: 0.40872883796691895
step: 180, loss: 0.29683220386505127
step: 190, loss: 0.16032689809799194
/usr/local/lib/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=0.46607669616519176, f1=0.44642857142857145,
best_f1=0.44642857142857145
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2929524779319763
```

```
step: 10, loss: 0.03338158503174782
step: 20, loss: 0.09596224874258041
step: 30, loss: 0.19741125404834747
step: 40, loss: 0.48275476694107056
step: 50, loss: 0.2826204299926758
step: 60, loss: 0.17881637811660767
step: 70, loss: 0.2030232846736908
step: 80, loss: 0.14457711577415466
step: 90, loss: 0.2023307979106903
step: 100, loss: 0.19528931379318237
step: 110, loss: 0.2497977316379547
step: 120, loss: 0.3456871211528778
step: 130, loss: 0.18758489191532135
step: 140, loss: 0.3711357116699219
step: 150, loss: 0.08208344131708145
step: 160, loss: 0.1594131737947464
step: 170, loss: 0.40555474162101746
step: 180, loss: 0.18158946931362152
step: 190, loss: 0.3806755542755127
/usr/local/lib/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=0.6887755102040816, f1=0.6851385390428211,
best_f1=0.6851385390428211
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.18198277056217194
step: 10, loss: 0.24562516808509827
step: 20, loss: 0.03024105727672577
step: 30, loss: 0.040213149040937424
step: 40, loss: 0.10798300802707672
step: 50, loss: 0.09596499055624008
step: 60, loss: 0.06858957558870316
step: 70, loss: 0.24957801401615143
step: 80, loss: 0.19484186172485352
step: 90, loss: 0.05160617455840111
step: 100, loss: 0.17634990811347961
step: 110, loss: 0.29813483357429504
step: 120, loss: 0.05362170562148094
step: 130, loss: 0.014036871492862701
step: 140, loss: 0.2235841006040573
step: 150, loss: 0.13746429979801178
step: 160, loss: 0.15505777299404144
```

```
step: 170, loss: 0.08898341655731201
step: 180, loss: 0.12184399366378784
step: 190, loss: 0.296196311712265
/usr/local/lib/python3.7/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=0.6756756756756755, f1=0.6869806094182824,
best_f1=0.6851385390428211
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08030256628990173
step: 10, loss: 0.112972192466259
step: 20, loss: 0.03520125150680542
step: 30, loss: 0.06889086961746216
step: 40, loss: 0.01866157352924347
step: 50, loss: 0.11849835515022278
step: 60, loss: 0.0540640652179718
step: 70, loss: 0.00807055551558733
step: 80, loss: 0.13333959877490997
step: 90, loss: 0.09020942449569702
step: 100, loss: 0.04741621017456055
step: 110, loss: 0.17505383491516113
step: 120, loss: 0.1322336196899414
step: 130, loss: 0.14680026471614838
step: 140, loss: 0.24969813227653503
step: 150, loss: 0.04344376176595688
step: 160, loss: 0.08656367659568787
step: 170, loss: 0.049486372619867325
step: 180, loss: 0.029279150068759918
step: 190, loss: 0.030873417854309082
/usr/local/lib/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=0.7394957983193278, f1=0.7570621468926554,
best_f1=0.7570621468926554
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02962687239050865
step: 10, loss: 0.015201007016003132
step: 20, loss: 0.057462360709905624
step: 30, loss: 0.006080050021409988
step: 40, loss: 0.040348730981349945
```

```
step: 50, loss: 0.23095746338367462
step: 60, loss: 0.01221542339771986
step: 70, loss: 0.013690273277461529
step: 80, loss: 0.009432127699255943
step: 90, loss: 0.028031444177031517
step: 100, loss: 0.01510606985539198
step: 110, loss: 0.031073568388819695
step: 120, loss: 0.002388094086199999
step: 130, loss: 0.0499250628054142
step: 140, loss: 0.004723282530903816
step: 150, loss: 0.022283386439085007
step: 160, loss: 0.027859121561050415
step: 170, loss: 0.021117620170116425
step: 180, loss: 0.014834470115602016
step: 190, loss: 0.14805862307548523
/usr/local/lib/python3.7/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=0.7287671232876712, f1=0.7401129943502824,
best_f1=0.7570621468926554
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.015247436240315437
step: 10, loss: 0.002449555555358529
step: 20, loss: 0.009459571912884712
step: 30, loss: 0.07883941382169724
step: 40, loss: 0.035573143512010574
step: 50, loss: 0.002863406203687191
step: 60, loss: 0.0010418544989079237
step: 70, loss: 0.007604395039379597
step: 80, loss: 0.0038354541175067425
step: 90, loss: 0.008564746007323265
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 100, loss: 0.002556223887950182
step: 110, loss: 0.1005745381116867
step: 120, loss: 0.007923168130218983
step: 130, loss: 0.012227599509060383
step: 140, loss: 0.044581376016139984
step: 150, loss: 0.04061779007315636
step: 160, loss: 0.008737461641430855
step: 170, loss: 0.003198132151737809
step: 180, loss: 0.005369626916944981
step: 190, loss: 0.008232822641730309
/usr/local/lib/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=0.7466666666666666, f1=0.7382920110192837,
best_f1=0.7382920110192837
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005575879476964474
step: 10, loss: 0.06452945619821548
step: 20, loss: 0.0019148363498970866
step: 30, loss: 0.003408627351745963
step: 40, loss: 0.0017343501094728708
step: 50, loss: 0.019614117220044136
step: 60, loss: 0.005055124405771494
step: 70, loss: 0.0008362550870515406
step: 80, loss: 0.10689535737037659
step: 90, loss: 0.0013905455125495791
step: 100, loss: 0.0048832399770617485
step: 110, loss: 0.0037929590325802565
step: 120, loss: 0.0008615341503173113
step: 130, loss: 0.00310566951520741
step: 140, loss: 0.018750814720988274
step: 150, loss: 0.13435310125350952
step: 160, loss: 0.14202678203582764
step: 170, loss: 0.0303681381046772
step: 180, loss: 0.05371781438589096
step: 190, loss: 0.02919442392885685
/usr/local/lib/python3.7/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=0.7258485639686684, f1=0.7282913165266107,
best_f1=0.7382920110192837
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04410267993807793
step: 10, loss: 0.0018289712024852633
step: 20, loss: 0.0008801182266324759
step: 30, loss: 0.0011837235651910305
step: 40, loss: 0.004667059518396854
step: 50, loss: 0.0011678938753902912
step: 60, loss: 0.00047637338866479695
step: 70, loss: 0.1003463864326477
step: 80, loss: 0.005552257876843214
step: 90, loss: 0.029183151200413704
step: 100, loss: 0.0024894699454307556
step: 110, loss: 0.0017817409243434668

```
step: 120, loss: 0.002220110734924674
step: 130, loss: 0.0026734662242233753
step: 140, loss: 0.0027100478764623404
step: 150, loss: 0.002146783284842968
step: 160, loss: 0.0019079847261309624
step: 170, loss: 0.0003312284534331411
step: 180, loss: 0.0012188247637823224
step: 190, loss: 0.08415098488330841
/usr/local/lib/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=0.7631578947368421, f1=0.7446808510638299,
best_f1=0.7446808510638299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0012074830010533333
step: 10, loss: 0.00087686261394992447
step: 20, loss: 0.008024855516850948
step: 30, loss: 0.007743341848254204
step: 40, loss: 0.010013674385845661
step: 50, loss: 0.004992590751498938
step: 60, loss: 0.001959201879799366
step: 70, loss: 0.00086536246817559
step: 80, loss: 0.011450227349996567
step: 90, loss: 0.006135708652436733
step: 100, loss: 0.0003346211451571435
step: 110, loss: 0.016555171625316143
step: 120, loss: 0.021694371476769447
step: 130, loss: 0.00041832009446807206
step: 140, loss: 0.0002577774866949767
step: 150, loss: 0.0008210456580854952
step: 160, loss: 0.0007759936270304024
step: 170, loss: 0.0030000482220202684
step: 180, loss: 0.0007967049605213106
step: 190, loss: 0.0007106431294232607
/usr/local/lib/python3.7/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=0.7534626038781163, f1=0.7283236994219653,
best_f1=0.7446808510638299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.0005833299946971238
step: 10, loss: 0.0013589084846898913
step: 20, loss: 0.00023089365276973695
step: 30, loss: 0.00041646271711215377
step: 40, loss: 0.0005108599434606731
step: 50, loss: 0.00039365599513985217
step: 60, loss: 0.0011693225242197514
step: 70, loss: 0.0007585844723507762
step: 80, loss: 0.0002617575810290873
step: 90, loss: 0.00046873881365172565
step: 100, loss: 0.02462589740753174
step: 110, loss: 0.0009187896503135562
step: 120, loss: 0.002104537794366479
step: 130, loss: 0.00049576599849388
step: 140, loss: 0.00044140798854641616
step: 150, loss: 0.00017295530415140092
step: 160, loss: 0.002144442405551672
step: 170, loss: 0.0004511508741416037
step: 180, loss: 0.013126925565302372
step: 190, loss: 0.005616732407361269
/usr/local/lib/python3.7/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=0.7582697201017812, f1=0.7139240506329113,
best_f1=0.7446808510638299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00017736364679876715
step: 10, loss: 0.17119188606739044
step: 20, loss: 0.005182386375963688
step: 30, loss: 0.22534029185771942
step: 40, loss: 0.00032808739342726767
step: 50, loss: 0.00027893518563359976
step: 60, loss: 0.0019874584395438433
step: 70, loss: 0.00021118843869771808
step: 80, loss: 0.00021341421233955577
step: 90, loss: 0.00045746820978820324
step: 100, loss: 0.00014450427261181176
step: 110, loss: 0.00012320144742261618
step: 120, loss: 0.00020628572383429855
step: 130, loss: 0.006764700636267662
step: 140, loss: 0.0002100891142617911
step: 150, loss: 0.00017483592091593891
step: 160, loss: 0.0001237048563780263
step: 170, loss: 0.0015585148939862847
step: 180, loss: 0.08407895267009735
step: 190, loss: 0.000168861573911272
```

```
/usr/local/lib/python3.7/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=0.7601078167115903, f1=0.7380281690140845,
best_f1=0.7446808510638299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000392377027310431
step: 10, loss: 0.00032634378294460475
step: 20, loss: 0.0002812123275361955
step: 30, loss: 0.0003506412322167307
step: 40, loss: 0.002463867422193289
step: 50, loss: 0.0014942586421966553
step: 60, loss: 0.00011763503425754607
step: 70, loss: 0.016426289454102516
step: 80, loss: 0.0014015280175954103
step: 90, loss: 0.00025811957311816514
step: 100, loss: 0.0002242983173346147
step: 110, loss: 0.00017548890900798142
step: 120, loss: 0.000393608002923429
step: 130, loss: 0.018696531653404236
step: 140, loss: 0.0004513626918196678
step: 150, loss: 0.0007336662383750081
step: 160, loss: 0.10172964632511139
step: 170, loss: 0.0007064375677146018
step: 180, loss: 0.004285057540982962
step: 190, loss: 0.08618733286857605
/usr/local/lib/python3.7/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=0.7474226804123711, f1=0.6894736842105263,
best_f1=0.7446808510638299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00021616432059090585
step: 10, loss: 0.0002875989302992821
step: 20, loss: 0.0029409106355160475
step: 30, loss: 0.0006583961076103151
step: 40, loss: 0.00019860915199387819
step: 50, loss: 0.0011576834367588162
step: 60, loss: 0.12132178992033005
step: 70, loss: 0.0010691058123484254
step: 80, loss: 0.0005015954957343638
step: 90, loss: 0.008172261528670788
step: 100, loss: 0.0003203694068361074
step: 110, loss: 0.0008024502894841135
```

```
step: 120, loss: 0.0003608637198340148
step: 130, loss: 0.00041528386645950377
step: 140, loss: 0.00021954749536234885
step: 150, loss: 0.022059958428144455
step: 160, loss: 0.00023070434690453112
step: 170, loss: 9.903834870783612e-05
step: 180, loss: 0.002306752372533083
step: 190, loss: 0.002620298881083727
/usr/local/lib/python3.7/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=0.7596899224806202, f1=0.7336956521739131,
best_f1=0.7446808510638299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00020216903067193925
step: 10, loss: 0.18171446025371552
step: 20, loss: 0.018644705414772034
step: 30, loss: 0.00018798747623804957
step: 40, loss: 0.0006468945648521185
step: 50, loss: 0.00871792808175087
step: 60, loss: 0.003233496565371752
step: 70, loss: 0.0008173327660188079
step: 80, loss: 0.00018529528460931033
step: 90, loss: 0.0007344318437390029
step: 100, loss: 0.00037060430622659624
step: 110, loss: 0.0004582289548125118
step: 120, loss: 0.0003574866568669677
step: 130, loss: 0.0002725672966334969
step: 140, loss: 0.0011874830815941095
step: 150, loss: 0.0006131020491011441
step: 160, loss: 0.0010833428241312504
step: 170, loss: 0.0008733998402021825
step: 180, loss: 0.00023690774105489254
step: 190, loss: 0.00020079004752915353
/usr/local/lib/python3.7/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=0.7500000000000001, f1=0.7252124645892353,
best_f1=0.7446808510638299
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00846523605287075
step: 10, loss: 0.004360022023320198
step: 20, loss: 0.00019061268540099263
step: 30, loss: 0.0009062004974111915
```

```
step: 40, loss: 0.0007816830184310675
step: 50, loss: 0.0004256390093360096
step: 60, loss: 0.00023646077897865325
step: 70, loss: 0.0006501839379779994
step: 80, loss: 0.00045909604523330927
step: 90, loss: 0.01877480000257492
step: 100, loss: 0.001289099338464439
step: 110, loss: 0.0006329793250188231
step: 120, loss: 0.001638914691284299
step: 130, loss: 0.00048436364158988
step: 140, loss: 0.00135442812461406
step: 150, loss: 0.015848658978939056
step: 160, loss: 0.0021858224645256996
step: 170, loss: 0.0006092911935411394
step: 180, loss: 0.00023347255773842335
step: 190, loss: 0.0011273128911852837
/usr/local/lib/python3.7/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=0.7520435967302452, f1=0.7252124645892353,
best_f1=0.7446808510638299
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.")
2049it [00:09, 223.34it/s]
load_f1 = 0.6980609418282547
real_f1 = 0.6881720430107526
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, 253.36it/s]
```

3.2 DK TEXTUAL

3.2.1 Abt-Buy - Running the matcher

```python
[ ]:  # 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 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: 0.8447359204292297
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.2185119241476059
step: 20, loss: 0.14797595143318176
step: 30, loss: 0.254518061876297
step: 40, loss: 0.3096427619457245
step: 50, loss: 0.3834676146507263
step: 60, loss: 0.4480210542678833
step: 70, loss: 0.30818867683410645
step: 80, loss: 0.25844940543174744
step: 90, loss: 0.41776207089424133
step: 100, loss: 0.2432185709476471
```

```
step: 110, loss: 0.18405435979366302
step: 120, loss: 0.5170094966888428
step: 130, loss: 0.41420212388038635
step: 140, loss: 0.4833517372608185
step: 150, loss: 0.09051317721605301
step: 160, loss: 0.34557321667671204
step: 170, loss: 0.24815770983695984
/usr/local/lib/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=0.2776349614395887, f1=0.26450742240215924,
best_f1=0.26450742240215924
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.4187040328979492
step: 10, loss: 0.2355216145515442
step: 20, loss: 0.3476466238498688
step: 30, loss: 0.27740219235420227
step: 40, loss: 0.21709808707237244
step: 50, loss: 0.266499787569046
step: 60, loss: 0.0913069024682045
step: 70, loss: 0.29338905215263367
step: 80, loss: 0.1806657612323761
step: 90, loss: 0.20741334557533264
step: 100, loss: 0.16699238121509552
step: 110, loss: 0.2921850383281708
step: 120, loss: 0.10747256129980087
step: 130, loss: 0.08211599290370941
step: 140, loss: 0.2411159873008728
step: 150, loss: 0.16573230922222137
step: 160, loss: 0.1318647861480713
step: 170, loss: 0.15133054554462433
/usr/local/lib/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=0.6807511737089201, f1=0.6477541371158393,
best_f1=0.6477541371158393
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08227738738059998
step: 10, loss: 0.11967489123344421
step: 20, loss: 0.035037148743867874
step: 30, loss: 0.19234313070774078
step: 40, loss: 0.015792574733495712
step: 50, loss: 0.18659977614879608
step: 60, loss: 0.21205027401447296
step: 70, loss: 0.0787908211350441
step: 80, loss: 0.11277937144041061
step: 90, loss: 0.13038720190525055
step: 100, loss: 0.06557846814393997
step: 110, loss: 0.030648859217762947
step: 120, loss: 0.08743561059236526
step: 130, loss: 0.1642531305551529
step: 140, loss: 0.012613757513463497
step: 150, loss: 0.04712966829538345
step: 160, loss: 0.07696522772312164
step: 170, loss: 0.14678062498569489
/usr/local/lib/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=0.7358916478555305, f1=0.688034188034188,
best_f1=0.688034188034188
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05724490061402321
step: 10, loss: 0.11313051730394363
step: 20, loss: 0.16729985177516937
step: 30, loss: 0.1450107842683792
step: 40, loss: 0.0802384614944458
step: 50, loss: 0.030629783868789673
step: 60, loss: 0.20966660976409912
step: 70, loss: 0.010577344335615635
step: 80, loss: 0.058273326605558395
step: 90, loss: 0.06704433262348175
step: 100, loss: 0.08968811482191086
step: 110, loss: 0.057265400886535645
step: 120, loss: 0.15664426982402802
step: 130, loss: 0.07723170518875122
step: 140, loss: 0.016189904883503914
step: 150, loss: 0.06400885432958603
step: 160, loss: 0.07886112481355667
```

```
step: 170, loss: 0.025600584223866463
/usr/local/lib/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=0.7487922705314008, f1=0.6713615023474179,
best_f1=0.6713615023474179
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03159167617559433
step: 10, loss: 0.07198892533779144
step: 20, loss: 0.021870525553822517
step: 30, loss: 0.09206955134868622
step: 40, loss: 0.014060646295547485
step: 50, loss: 0.0550447553396225
step: 60, loss: 0.026114333420991898
step: 70, loss: 0.05247151479125023
step: 80, loss: 0.017189528793096542
step: 90, loss: 0.012787071987986565
step: 100, loss: 0.043214816600084305
step: 110, loss: 0.1063096672296524
step: 120, loss: 0.066630115956068039
step: 130, loss: 0.020469743758440018
step: 140, loss: 0.14778928458690643
step: 150, loss: 0.0369558110833168
step: 160, loss: 0.0588616281747818
step: 170, loss: 0.06312073767185211
/usr/local/lib/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=0.7518796992481204, f1=0.6960784313725491,
best_f1=0.6960784313725491
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01402944978326559
step: 10, loss: 0.17978280782699585
step: 20, loss: 0.042615655809640884
step: 30, loss: 0.02655099518597126
step: 40, loss: 0.009567268192768097
```

```
step: 50, loss: 0.05693545192480087
step: 60, loss: 0.005065268836915493
step: 70, loss: 0.0355156846344471
step: 80, loss: 0.02570047415792942
step: 90, loss: 0.013395008631050587
step: 100, loss: 0.048367343842983246
step: 110, loss: 0.16989922523498535
step: 120, loss: 0.09450266510248184
step: 130, loss: 0.1634443998336792
step: 140, loss: 0.030086426064372063
step: 150, loss: 0.27986854314804077
step: 160, loss: 0.0314653106033802
step: 170, loss: 0.007258689031004906
/usr/local/lib/python3.7/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=0.7335092348284961, f1=0.6917293233082706,
best_f1=0.6960784313725491
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005253925919532776
step: 10, loss: 0.00165498291160794616
step: 20, loss: 0.009080762974917889
step: 30, loss: 0.23545366525650024
step: 40, loss: 0.017851104959845543
step: 50, loss: 0.004385174252092838
step: 60, loss: 0.2100452333688736
step: 70, loss: 0.025138461962342262
step: 80, loss: 0.0035359819885343313
step: 90, loss: 0.04058254882693291
step: 100, loss: 0.014467120170593262
step: 110, loss: 0.001643993309698999
step: 120, loss: 0.2358594834804535
step: 130, loss: 0.03332772105932236
step: 140, loss: 0.01614160649478435
step: 150, loss: 0.016263503581285477
step: 160, loss: 0.003666343865916133
step: 170, loss: 0.2586556375026703
/usr/local/lib/python3.7/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=0.7146529562982005, f1=0.6748166259168704,
best_f1=0.6960784313725491
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08883526921272278
```

```
step: 10, loss: 0.0023299623280763626
step: 20, loss: 0.014226030558347702
step: 30, loss: 0.010798102244734764
step: 40, loss: 0.002858178224414587
step: 50, loss: 0.001798250712454319
step: 60, loss: 0.009427161887288094
step: 70, loss: 0.028828119859099388
step: 80, loss: 0.0012873687082901597
step: 90, loss: 0.012660348787903786
step: 100, loss: 0.024538137018680573
step: 110, loss: 0.06472413241863251
step: 120, loss: 0.003783557331189513
step: 130, loss: 0.04616234451532364
step: 140, loss: 0.0031982585787773132
step: 150, loss: 0.05931400880217552
step: 160, loss: 0.10305511951446533
step: 170, loss: 0.009534968994557858
/usr/local/lib/python3.7/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=0.7466666666666667, f1=0.7080103359173127,
best_f1=0.6960784313725491
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005689740180969238
step: 10, loss: 0.008867527358233929
step: 20, loss: 0.0006725778221152723
step: 30, loss: 0.009344722144305706
step: 40, loss: 0.009246100671589375
step: 50, loss: 0.056673843413591385
step: 60, loss: 0.0008362063090316951
step: 70, loss: 0.027615156024694443
step: 80, loss: 0.00181251869071275
step: 90, loss: 0.008497681468725204
step: 100, loss: 0.009970703162252903
step: 110, loss: 0.00024480969295836985
step: 120, loss: 0.02205859310925007
step: 130, loss: 0.0024690169375389814
step: 140, loss: 0.004968215245753527
step: 150, loss: 0.008532688021659851
step: 160, loss: 0.057234469801187515
step: 170, loss: 0.03261983394622803
/usr/local/lib/python3.7/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=0.7287234042553191, f1=0.695214105793451,
best_f1=0.6960784313725491
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0011591549264267087
step: 10, loss: 0.02058831974864006
step: 20, loss: 0.012613259255886078
step: 30, loss: 0.0005742716602981091
step: 40, loss: 0.0015140559989959002
step: 50, loss: 0.00189330021929115057
step: 60, loss: 0.001786844339221716
step: 70, loss: 0.002769528655335307
step: 80, loss: 0.001460416940972209
step: 90, loss: 0.0850272923707962
step: 100, loss: 0.06370753049850464
step: 110, loss: 0.002375712152570486
step: 120, loss: 0.04404887557029724
step: 130, loss: 0.0294588228714466
step: 140, loss: 0.01383874099701643
step: 150, loss: 0.00671102665647423
step: 160, loss: 0.000524491595569998
step: 170, loss: 0.001337959198281169
/usr/local/lib/python3.7/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=0.7307692307692308, f1=0.6824146981627297,
best_f1=0.6960784313725491
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006746025756001472
step: 10, loss: 0.042617328464984894
step: 20, loss: 0.06779395788908005
step: 30, loss: 0.0027401375118643045
step: 40, loss: 0.018489159643650055
step: 50, loss: 0.002013358986005187
step: 60, loss: 0.034197140485048294
step: 70, loss: 0.004388872999697924
step: 80, loss: 0.0009289636509492993
step: 90, loss: 0.002460867166519165
step: 100, loss: 0.0003470202209427953
step: 110, loss: 0.0005117470864206553
step: 120, loss: 0.01779281534254551
step: 130, loss: 0.013793845660984516
step: 140, loss: 0.028088729828596115
step: 150, loss: 0.004704730585217476
step: 160, loss: 0.0015957854920998216
step: 170, loss: 0.007745022885501385
/usr/local/lib/python3.7/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=0.7262872628726287, f1=0.690537084398977,
best_f1=0.6960784313725491
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00042614215635694563
step: 10, loss: 0.0016734987730160356
step: 20, loss: 0.0025803218595683575
step: 30, loss: 0.00033528235508129
step: 40, loss: 0.0005769909475930035
step: 50, loss: 0.001741513260640204
step: 60, loss: 0.0003951621474698186
step: 70, loss: 0.004217409063130617
step: 80, loss: 0.009365440346300602
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 90, loss: 0.00022400241869036108
step: 100, loss: 0.00044264638563618064
step: 110, loss: 0.0006552557460963726
step: 120, loss: 0.005560798104852438
step: 130, loss: 0.007436810992658138
step: 140, loss: 0.0007070120191201568
step: 150, loss: 0.006435503251850605
step: 160, loss: 0.000930282985791564
step: 170, loss: 0.003183159278705716
/usr/local/lib/python3.7/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=0.723514211886305, f1=0.6828087167070218,
best_f1=0.6960784313725491
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003580857301130891
step: 10, loss: 0.04988064244389534
step: 20, loss: 0.01530646625906229
step: 30, loss: 0.005301825236529112
step: 40, loss: 0.01083078607916832
step: 50, loss: 0.0009287673747166991
step: 60, loss: 0.010526907630264759
step: 70, loss: 0.0030237010214477777
step: 80, loss: 0.0005208681104704738
step: 90, loss: 0.0024014124646782875
step: 100, loss: 0.006885797716677189
step: 110, loss: 0.01801201142370701
step: 120, loss: 0.015741486102342606
step: 130, loss: 0.015688277781009674
```

```
step: 140, loss: 0.0001876793976407498
step: 150, loss: 0.013049378991127014
step: 160, loss: 0.0009498127619735897
step: 170, loss: 0.027273530140519142
/usr/local/lib/python3.7/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=0.7305699481865284, f1=0.6891566265060242,
best_f1=0.6960784313725491
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008709542453289032
step: 10, loss: 0.00020167624461464584
step: 20, loss: 0.0040599144995212555
step: 30, loss: 0.0010376506252214313
step: 40, loss: 0.0009931594831869006
step: 50, loss: 0.0028076861053705215
step: 60, loss: 0.0007085507386364043
step: 70, loss: 0.028666799888014793
step: 80, loss: 0.0004000875633209944
step: 90, loss: 0.00037169564166106284
step: 100, loss: 0.0003493308904580772
step: 110, loss: 0.0002658023440744728
step: 120, loss: 0.00030900101410225034
step: 130, loss: 0.0676349326968193
step: 140, loss: 0.025249596685171127
step: 150, loss: 0.010293868370354176
step: 160, loss: 0.08328524231910706
step: 170, loss: 0.00029857884510420263
/usr/local/lib/python3.7/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=0.7395833333333331, f1=0.6909975669099757,
best_f1=0.6960784313725491
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005628286395221949
step: 10, loss: 0.0022449863608926535
step: 20, loss: 0.010442489758133888
step: 30, loss: 0.08892887085676193
step: 40, loss: 0.00022726744646206498
step: 50, loss: 0.0003986226220149547
step: 60, loss: 0.00015881711442489177
step: 70, loss: 0.0012359301326796412
step: 80, loss: 0.007874148897826672
step: 90, loss: 0.00803446676582098
```

```
step: 100, loss: 0.0005448363372124732
step: 110, loss: 0.08590254932641983
step: 120, loss: 0.005987165030092001
step: 130, loss: 0.0009821916464716196
step: 140, loss: 0.09344137459993362
step: 150, loss: 0.01716078445315361
step: 160, loss: 0.0006809187470935285
step: 170, loss: 0.0003141153429169208
/usr/local/lib/python3.7/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=0.7379134860050892, f1=0.6886792452830189,
best_f1=0.6960784313725491
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 [00:05, 322.28it/s]
load_f1 = 0.7227722772277227
real_f1 = 0.7073170731707317
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, 252.09it/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/
```

```
Downloading: 100% 442/442 [00:00<00:00, 507kB/s]
Downloading: 100% 232k/232k [00:00<00:00, 2.50MB/s]
Downloading: 100% 268M/268M [00:03<00:00, 67.8MB/s]
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: 0.7977885007858276
step: 10, loss: 0.47171780467033386
step: 20, loss: 0.5826108455657959
step: 30, loss: 0.4938201606273651
step: 40, loss: 0.3953050971031189
step: 50, loss: 0.35717830061912537
step: 60, loss: 0.25754547119140625
step: 70, loss: 0.16313371062278748
step: 80, loss: 0.32194533944129944
step: 90, loss: 0.08157812803983688
step: 100, loss: 0.1351412683725357
step: 110, loss: 0.1466190665960312
step: 120, loss: 0.041793111711740494
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 130, loss: 0.0461166575551033
step: 140, loss: 0.09841186553239822
```

```
step: 150, loss: 0.1056588813662529
step: 160, loss: 0.18748660385608673
step: 170, loss: 0.025558898225426674
step: 180, loss: 0.007389332167804241
step: 190, loss: 0.3163536489009857
step: 200, loss: 0.04778594523668289
step: 210, loss: 0.05420665815472603
step: 220, loss: 0.04983772709965706
step: 230, loss: 0.019763000309467316
/usr/local/lib/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=0.9371554575523705, f1=0.9289740698985345,
best_f1=0.9289740698985345
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11386702954769135
step: 10, loss: 0.10373307019472122
step: 20, loss: 0.02531144954264164
step: 30, loss: 0.06679247319698334
step: 40, loss: 0.055001188069581985
step: 50, loss: 0.03589959070086479
step: 60, loss: 0.021723132580518723
step: 70, loss: 0.10690712183713913
step: 80, loss: 0.0040284874849021435
step: 90, loss: 0.01786421798169613
step: 100, loss: 0.15191468596458435
step: 110, loss: 0.06271378695964813
step: 120, loss: 0.06387464702129364
step: 130, loss: 0.13909219205379486
step: 140, loss: 0.16073231399059296
step: 150, loss: 0.0403754860162735
step: 160, loss: 0.14100955426692963
step: 170, loss: 0.052075427025556564
step: 180, loss: 0.01155760232359171
step: 190, loss: 0.12707985937595367
step: 200, loss: 0.1183665543794632
step: 210, loss: 0.11741166561841965
step: 220, loss: 0.003219516947865486
step: 230, loss: 0.08089778572320938
/usr/local/lib/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=0.9384098544232924, f1=0.9465301478953355,
best_f1=0.9465301478953355
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08701024204492569
step: 10, loss: 0.05237084999680519
step: 20, loss: 0.03297644481062889
step: 30, loss: 0.029846027493476868
step: 40, loss: 0.16685399413108826
step: 50, loss: 0.04648960381746292
step: 60, loss: 0.09414080530405045
step: 70, loss: 0.008882452733814716
step: 80, loss: 0.01975242793560028
step: 90, loss: 0.12438483536243439
step: 100, loss: 0.012947270646691322
step: 110, loss: 0.04492582380771637
step: 120, loss: 0.013936996459960938
step: 130, loss: 0.0037605317775160074
step: 140, loss: 0.0017200721194967628
step: 150, loss: 0.001639602705836296
step: 160, loss: 0.004795771092176437
step: 170, loss: 0.14416088163852692
step: 180, loss: 0.009805064648389816
step: 190, loss: 0.05256135016679764
step: 200, loss: 0.0406491793692112
step: 210, loss: 0.012876053340733051
step: 220, loss: 0.009234538301825523
step: 230, loss: 0.07237989455461502
/usr/local/lib/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=0.9598214285714285, f1=0.9479638009049774,
best_f1=0.9479638009049774
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.030396828427910805
step: 10, loss: 0.008617435581982136
step: 20, loss: 0.06928694248199463
```

```
step: 30, loss: 0.002149739535525441
step: 40, loss: 0.017606167122721672
step: 50, loss: 0.003987040836364031
step: 60, loss: 0.006786162964999676
step: 70, loss: 0.008155610412359238
step: 80, loss: 0.10718601197004318
step: 90, loss: 0.08000728487968445
step: 100, loss: 0.0046770297922194
step: 110, loss: 0.02410355769097805
step: 120, loss: 0.013189934194087982
step: 130, loss: 0.053980324417352676
step: 140, loss: 0.018356595188379288
step: 150, loss: 0.005098107270896435
step: 160, loss: 0.0018948308425024152
step: 170, loss: 0.006462803576141596
step: 180, loss: 0.04665207862854004
step: 190, loss: 0.002869817428290844
step: 200, loss: 0.0027161238249391317
step: 210, loss: 0.01634102687239647
step: 220, loss: 0.004113192670047283
step: 230, loss: 0.0024936676491051912
/usr/local/lib/python3.7/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=0.9444444444444444, f1=0.9367369589345171,
best_f1=0.9479638009049774
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04101866856217384
step: 10, loss: 0.06464467942714691
step: 20, loss: 0.022314608097076416
step: 30, loss: 0.00075122358975932
step: 40, loss: 0.008395036682486534
step: 50, loss: 0.0018497987184673548
step: 60, loss: 0.0035686874762177467
step: 70, loss: 0.0006946175126358867
step: 80, loss: 0.0013808173825964332
step: 90, loss: 0.0007994745392352343
step: 100, loss: 0.02149426005780697
step: 110, loss: 0.0019644771236926317
step: 120, loss: 0.04774923622608185
step: 130, loss: 0.010034704580903053
step: 140, loss: 0.0015676917973905802
step: 150, loss: 0.003376706736162305
step: 160, loss: 0.1659112274646759
step: 170, loss: 0.008413535542786121
step: 180, loss: 0.012881215661764145
```

```
step: 190, loss: 0.0027898431289941072
step: 200, loss: 0.0031790430657565594
step: 210, loss: 0.0018737869104370475
step: 220, loss: 0.001263050944544375
step: 230, loss: 0.0010744382161647081
/usr/local/lib/python3.7/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=0.9487459105779716, f1=0.9525909592061743,
best_f1=0.9479638009049774
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0028881325852125883
step: 10, loss: 0.029838742688298225
step: 20, loss: 0.005045570433139801
step: 30, loss: 0.0015460718423128128
step: 40, loss: 0.008475571870803833
step: 50, loss: 0.006190781015902758
step: 60, loss: 0.005852648988366127
step: 70, loss: 0.0017503689741715789
step: 80, loss: 0.00139673484954983
step: 90, loss: 0.0006047565257176757
step: 100, loss: 0.00649120332673192
step: 110, loss: 0.08214198052883148
step: 120, loss: 0.0023427328560501337
step: 130, loss: 0.001928898273035884
step: 140, loss: 0.007581348996609449
step: 150, loss: 0.00091649399837479
step: 160, loss: 0.049200110137462616
step: 170, loss: 0.004178515635430813
step: 180, loss: 0.006171730812638998
step: 190, loss: 0.00979720987379551
step: 200, loss: 0.026249542832374573
step: 210, loss: 0.06296635419130325
step: 220, loss: 0.0014455930795520544
step: 230, loss: 0.009168514050543308
/usr/local/lib/python3.7/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=0.943646408839779, f1=0.9474860335195531,
best_f1=0.9479638009049774
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.19021227955818176
step: 10, loss: 0.002048509195446968
step: 20, loss: 0.001306361285969615
```

```
step: 30, loss: 0.0026178371626883745
step: 40, loss: 0.0026581797283142805
step: 50, loss: 0.0006841907161287963
step: 60, loss: 0.008219673298299313
step: 70, loss: 0.0009313377668149769
step: 80, loss: 0.0003656026383396238
step: 90, loss: 0.0017064183484762907
step: 100, loss: 0.0005026694270782173
step: 110, loss: 0.0007357572321780026
step: 120, loss: 0.0954316109418869
step: 130, loss: 0.02964283712208271
step: 140, loss: 0.0007044555386528373
step: 150, loss: 0.0005595532711595297
step: 160, loss: 0.000592812430113554
step: 170, loss: 0.001978813437744975
step: 180, loss: 0.0028487660456448793
step: 190, loss: 0.0020977642852813005
step: 200, loss: 0.006743014790117741
step: 210, loss: 0.0010687625035643578
step: 220, loss: 0.0009364539291709661
step: 230, loss: 0.03908173739910126
/usr/local/lib/python3.7/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=0.9511111111111112, f1=0.9457013574660633,
best_f1=0.9479638009049774
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00400164071470499
step: 10, loss: 0.019374750554561615
step: 20, loss: 0.0005522326100617647
step: 30, loss: 0.03702152892947197
step: 40, loss: 0.0009405020973645151
step: 50, loss: 0.010271935723721981
step: 60, loss: 0.00047609410830773413
step: 70, loss: 0.005429540295153856
step: 80, loss: 0.002202720148488879
step: 90, loss: 0.03860677778720856
step: 100, loss: 0.002140214666724205
step: 110, loss: 0.00045867369044572115
step: 120, loss: 0.00022198895749170333
step: 130, loss: 0.0005526302848011255
step: 140, loss: 0.00026993209030479193
step: 150, loss: 0.00021469854982569814
step: 160, loss: 0.0005251511465758085
step: 170, loss: 0.00049683987716300726
step: 180, loss: 0.004076933953911066
```

```
step: 190, loss: 0.0014458841178566217
step: 200, loss: 0.0007153634796850383
step: 210, loss: 0.00025114466552622616
step: 220, loss: 0.00017010040755849332
step: 230, loss: 0.22308027744293213
/usr/local/lib/python3.7/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=0.9482948294829483, f1=0.9485458612975392,
best_f1=0.9479638009049774
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04738464206457138
step: 10, loss: 0.014497748576104641
step: 20, loss: 0.012898939661681652
step: 30, loss: 0.0008291436824947596
step: 40, loss: 0.0005013038753531873
step: 50, loss: 0.07927379757165909
step: 60, loss: 0.006177426315844059
step: 70, loss: 0.0018577426671981812
step: 80, loss: 0.10376135259866714
step: 90, loss: 0.002037618076428771
step: 100, loss: 0.0003603090299293399
step: 110, loss: 0.00040055817225947976
step: 120, loss: 0.040650032460689545
step: 130, loss: 0.0003346760349813849
step: 140, loss: 0.0028034320566803217
step: 150, loss: 0.00015051252557896078
step: 160, loss: 0.0018238211050629616
step: 170, loss: 0.00047887087566778064
step: 180, loss: 0.019211186096072197
step: 190, loss: 0.0022947771940380335
step: 200, loss: 0.07304955273866653
step: 210, loss: 0.0004690026107709855
step: 220, loss: 0.00956203043460846
step: 230, loss: 0.0007235648226924241
/usr/local/lib/python3.7/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=0.9524861878453039, f1=0.9473684210526316,
best_f1=0.9479638009049774
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001125274458900094
step: 10, loss: 0.00016230846813414246
step: 20, loss: 0.0005988606717437506
```

```
step: 30, loss: 0.00023831722501199692
step: 40, loss: 0.00019796271226368845
step: 50, loss: 0.0012446132022887468
step: 60, loss: 0.09826718270778656
step: 70, loss: 0.0048406897112727165
step: 80, loss: 0.00027769707958213985
step: 90, loss: 0.0024116418790072203
step: 100, loss: 0.0010352286044508219
step: 110, loss: 0.00035468663554638624
step: 120, loss: 0.0101505471393466
step: 130, loss: 0.002519231056794524
step: 140, loss: 0.0032332511618733406
step: 150, loss: 0.00012443069135770202
step: 160, loss: 0.00014134046796243638
step: 170, loss: 0.0005444528069347143
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 180, loss: 0.13771778345108032
step: 190, loss: 0.000654245144687593
step: 200, loss: 0.00013092691369820386
step: 210, loss: 0.00024212485004682094
step: 220, loss: 0.0009639140916988254
step: 230, loss: 0.0007067658007144928
/usr/local/lib/python3.7/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=0.9470198675496688, f1=0.9453734671125975,
best_f1=0.9479638009049774
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001391887926729396
step: 10, loss: 0.0005239841993898153
step: 20, loss: 0.00017475760250817984
step: 30, loss: 0.00024347545695491135
step: 40, loss: 0.0003567553358152509
step: 50, loss: 0.0003501740866340697
step: 60, loss: 0.0012299712980166078
step: 70, loss: 0.000343008832913518
step: 80, loss: 0.0002568041963968426
step: 90, loss: 0.003922153729945421
step: 100, loss: 0.00030365094426088035
step: 110, loss: 0.0001438676263205707
step: 120, loss: 0.0006474334513768554
step: 130, loss: 0.00016004382632672787
step: 140, loss: 0.00016014421998988837
step: 150, loss: 0.0007599511882290244
step: 160, loss: 0.001476808451116085
step: 170, loss: 0.004366862587630749
```

```
step: 180, loss: 0.004316161852329969
step: 190, loss: 0.00027533958200365305
step: 200, loss: 0.0001891755819087848
step: 210, loss: 0.00016593930195085704
step: 220, loss: 0.0007811755058355629
step: 230, loss: 0.0018387967720627785
/usr/local/lib/python3.7/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=0.94211123723042, f1=0.9402298850574713,
best_f1=0.9479638009049774
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003084450727328658
step: 10, loss: 0.0004783452895935625
step: 20, loss: 0.00017138604016508907
step: 30, loss: 0.0004146458231844008
step: 40, loss: 0.00010302550799679011
step: 50, loss: 8.391863229917362e-05
step: 60, loss: 0.009438241831958294
step: 70, loss: 0.00016417927690781653
step: 80, loss: 0.0001222628925461322
step: 90, loss: 0.00032979692332446575
step: 100, loss: 7.80218470026739e-05
step: 110, loss: 0.00013163154653739184
step: 120, loss: 9.923594916472211e-05
step: 130, loss: 0.00027608725940808654
step: 140, loss: 0.0005769152194261551
step: 150, loss: 9.719667286844924e-05
step: 160, loss: 0.002140249591320753
step: 170, loss: 0.00022713886573910713
step: 180, loss: 0.00017172022489830852
step: 190, loss: 0.00012515424168668687
step: 200, loss: 0.0023237131536006927
step: 210, loss: 7.323625322896987e-05
step: 220, loss: 0.00036426677252165973
step: 230, loss: 0.00024391907209064811
/usr/local/lib/python3.7/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=0.9471264367816091, f1=0.9422632794457275,
best_f1=0.9479638009049774
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.1531037974637e-05
step: 10, loss: 7.89816549513489e-05
```

```
step: 20, loss: 0.00021227338584139943
step: 30, loss: 0.02020101249217987
step: 40, loss: 0.00010554712935118005
step: 50, loss: 0.00017312588170170784
step: 60, loss: 0.00020536998636089265
step: 70, loss: 0.00023272840189747512
step: 80, loss: 0.0011238069273531437
step: 90, loss: 5.4750486015109345e-05
step: 100, loss: 0.00015154051652643830
step: 110, loss: 9.636852337280288e-05
step: 120, loss: 0.00011714937136275694
step: 130, loss: 0.00011373918096069247
step: 140, loss: 0.007043068762868643
step: 150, loss: 0.0006405016174539924
step: 160, loss: 8.043631532927975e-05
step: 170, loss: 9.108307858696207e-05
step: 180, loss: 0.00011744484800146893
step: 190, loss: 0.0003757250669877976
step: 200, loss: 0.0005506645538844168
step: 210, loss: 0.00019003944180440158
step: 220, loss: 0.00023221566516440362
step: 230, loss: 6.594559090444818e-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=0.9502762430939226, f1=0.9508928571428572,
best_f1=0.9479638009049774
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00011694958084262908
step: 10, loss: 4.6740031393710524e-05
step: 20, loss: 0.00024412466154899448
step: 30, loss: 0.00018298046779818833
step: 40, loss: 9.870126086752862e-05
step: 50, loss: 0.00014911578909959644
step: 60, loss: 4.875754893873818e-05
step: 70, loss: 8.551972132408991e-05
step: 80, loss: 0.00016968310228548944
step: 90, loss: 0.0004304655594751239
step: 100, loss: 0.000215998647036039
step: 110, loss: 0.00015791891200933605
step: 120, loss: 6.340788240777329e-05
step: 130, loss: 0.000193584623048082
step: 140, loss: 8.618066931376234e-05
step: 150, loss: 9.678076457930729e-05
step: 160, loss: 6.186541577335447e-05
step: 170, loss: 0.00017048961308319122
```

```
step: 180, loss: 0.0001536565978312865
step: 190, loss: 0.00012340409739408642
step: 200, loss: 6.29520436632447e-05
step: 210, loss: 0.00278626405633986
step: 220, loss: 0.00011117068061139435
step: 230, loss: 6.257898348849267e-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=0.9501661129568106, f1=0.9519553072625698,
best_f1=0.9479638009049774
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.773292261641473e-05
step: 10, loss: 0.00017532311903778464
step: 20, loss: 0.00013777667481917888
step: 30, loss: 0.0004814527928829193
step: 40, loss: 0.00012524322664830834
step: 50, loss: 0.00010689692862797529
step: 60, loss: 4.1930103179765865e-05
step: 70, loss: 0.0001001342388917692
step: 80, loss: 7.40452524041757e-05
step: 90, loss: 4.614569843397476e-05
step: 100, loss: 8.688222442287952e-05
step: 110, loss: 0.00010922431101789698
step: 120, loss: 0.00020937631779816002
step: 130, loss: 0.00011841639206977561
step: 140, loss: 0.005842987447977066
step: 150, loss: 0.00011401013034628704
step: 160, loss: 6.196399772306904e-05
step: 170, loss: 5.445994975161739e-05
step: 180, loss: 8.378813799936324e-05
step: 190, loss: 0.0005321577191352844
step: 200, loss: 6.712995673296973e-05
step: 210, loss: 0.000438223680248484
step: 220, loss: 0.0001456660102121532
step: 230, loss: 8.161352889146656e-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=0.9501661129568106, f1=0.9519553072625698,
best_f1=0.9479638009049774
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 [00:07, 332.49it/s]
load_f1 = 0.9470198675496688
real_f1 = 0.947136563876652
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, 414.59it/s]
```

3.3.2 DBLP-GoogleScholar - Running the matcher

```python
[ ]:  # 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 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: 0.797282338142395
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4381572902202606
step: 20, loss: 0.5361637473106384
step: 30, loss: 0.470694899559021
step: 40, loss: 0.45946305990219116
step: 50, loss: 0.44026052951812744
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 60, loss: 0.41191378235816956
step: 70, loss: 0.17213290929794312
step: 80, loss: 0.54940265417099
step: 90, loss: 0.37176766991615295
step: 100, loss: 0.20167188346385956
step: 110, loss: 0.12563863396644592
step: 120, loss: 0.164622962474823
step: 130, loss: 0.055475566536188126
step: 140, loss: 0.41605865955352783
step: 150, loss: 0.06531757861375809
step: 160, loss: 0.15003541111946106
step: 170, loss: 0.3696750998497009
step: 180, loss: 0.18651746213436127
step: 190, loss: 0.06454052031040192
step: 200, loss: 0.21349412202835083
step: 210, loss: 0.13785040378570557
step: 220, loss: 0.06063498929142952
step: 230, loss: 0.08765969425439835
step: 240, loss: 0.20303472876548767
step: 250, loss: 0.09010292589664459
step: 260, loss: 0.010502072982490063
step: 270, loss: 0.016611667349934578
step: 280, loss: 0.19776740670204163
step: 290, loss: 0.07576479017734528
step: 300, loss: 0.12024270743131638
step: 310, loss: 0.0715879276394844
step: 320, loss: 0.10092749446630478
step: 330, loss: 0.24463598430156708
step: 340, loss: 0.21330855786800385
step: 350, loss: 0.10220780968666077
```

```
step: 360, loss: 0.06618399918079376
step: 370, loss: 0.16745974123477936
step: 380, loss: 0.23667371273040771
step: 390, loss: 0.02865820750594139
step: 400, loss: 0.03206267207860947
step: 410, loss: 0.09886012971401215
step: 420, loss: 0.07681727409362793
step: 430, loss: 0.22004497051239014
step: 440, loss: 0.14609883725643158
step: 450, loss: 0.08858579397201538
step: 460, loss: 0.09028096497058868
step: 470, loss: 0.1873289793729782
step: 480, loss: 0.3006719946861267
step: 490, loss: 0.11710362881422043
step: 500, loss: 0.036209359765052795
step: 510, loss: 0.07737740129232407
step: 520, loss: 0.12792637944221497
step: 530, loss: 0.07422596216201782
/usr/local/lib/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=0.9011709601873535, f1=0.8905519176800748,
best_f1=0.8905519176800748
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.15972748398780823
step: 10, loss: 0.22207625210285187
step: 20, loss: 0.13699108362197876
step: 30, loss: 0.10502754151821136
step: 40, loss: 0.009485970251262188
step: 50, loss: 0.032837480306625366
step: 60, loss: 0.17163223028182983
step: 70, loss: 0.23936857283115387
step: 80, loss: 0.061580169945955276
step: 90, loss: 0.17062625288963318
step: 100, loss: 0.21487727761268616
step: 110, loss: 0.05369604751467705
step: 120, loss: 0.053170155733823776
step: 130, loss: 0.09085476398468018
step: 140, loss: 0.03497575595974922
step: 150, loss: 0.037166088819503784
step: 160, loss: 0.06804423779249191
step: 170, loss: 0.15605483949184418
```

```
step: 180, loss: 0.04043520987033844
step: 190, loss: 0.031029194593429565
step: 200, loss: 0.025034798309206963
step: 210, loss: 0.01024729572236538
step: 220, loss: 0.1443796306848526
step: 230, loss: 0.059119634330272675
step: 240, loss: 0.15724530816078186
step: 250, loss: 0.12526986002922058
step: 260, loss: 0.05000067874789238
step: 270, loss: 0.08666227012872696
step: 280, loss: 0.1467241644859314
step: 290, loss: 0.09078375995159149
step: 300, loss: 0.020526554435491562
step: 310, loss: 0.11931157112121582
step: 320, loss: 0.14093336462974548
step: 330, loss: 0.05796022340655327
step: 340, loss: 0.006093531381338835
step: 350, loss: 0.13035422563552856
step: 360, loss: 0.13591714203357697
step: 370, loss: 0.09150693565607071
step: 380, loss: 0.1263812929391861
step: 390, loss: 0.04842739924788475
step: 400, loss: 0.11346912384033203
step: 410, loss: 0.029394280165433884
step: 420, loss: 0.13386259973049164
step: 430, loss: 0.04000530019402504
step: 440, loss: 0.011352281086146832
step: 450, loss: 0.046582356095314026
step: 460, loss: 0.22234812378883362
step: 470, loss: 0.06957436352968216
step: 480, loss: 0.3781655430793762
step: 490, loss: 0.039861928671598434
step: 500, loss: 0.03953102231025696
step: 510, loss: 0.05724671110510826
step: 520, loss: 0.050416942685842514
step: 530, loss: 0.13437232375144958
/usr/local/lib/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=0.9072450392247347, f1=0.8967114404817044,
best_f1=0.8967114404817044
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.031039293855428696
step: 10, loss: 0.04072072356939316
step: 20, loss: 0.15149356424808502
step: 30, loss: 0.06729619204998016
step: 40, loss: 0.005167340859770775
step: 50, loss: 0.021029511466622353
step: 60, loss: 0.003343126503750682
step: 70, loss: 0.054645687341690063
step: 80, loss: 0.11309605836868286
step: 90, loss: 0.22833406925201416
step: 100, loss: 0.04963088408112526
step: 110, loss: 0.05199692025780678
step: 120, loss: 0.029929358512163162
step: 130, loss: 0.039914797991514206
step: 140, loss: 0.05490243062376976
step: 150, loss: 0.0019938619807362556
step: 160, loss: 0.004732377361506224
step: 170, loss: 0.018921274691820145
step: 180, loss: 0.019439080730080605
step: 190, loss: 0.020347265526652336
step: 200, loss: 0.08001501113176346
step: 210, loss: 0.131010040640831
step: 220, loss: 0.014691818505525589
step: 230, loss: 0.060467615723609924
step: 240, loss: 0.20008960366249084
step: 250, loss: 0.03356989473104477
step: 260, loss: 0.008919358253479004
step: 270, loss: 0.10965513437986374
step: 280, loss: 0.07285510003566742
step: 290, loss: 0.10549629479646683
step: 300, loss: 0.10845974087715149
step: 310, loss: 0.0315048024058342
step: 320, loss: 0.1780410259962082
step: 330, loss: 0.012999670580029488
step: 340, loss: 0.0044219838455319405
step: 350, loss: 0.15708863735198975
step: 360, loss: 0.008732171729207039
step: 370, loss: 0.1449793577194214
step: 380, loss: 0.009962761774659157
step: 390, loss: 0.04834398627281189
step: 400, loss: 0.04711716249585152
step: 410, loss: 0.11509226262569427
step: 420, loss: 0.11855288594961166
step: 430, loss: 0.026413410902023315
step: 440, loss: 0.024699682369828224
step: 450, loss: 0.18802586197853088
step: 460, loss: 0.1604214459657669
step: 470, loss: 0.025013813748955727
```

```
step: 480, loss: 0.019723286852240562
step: 490, loss: 0.006496070884168148
step: 500, loss: 0.04881182312965393
step: 510, loss: 0.11258653551340103
step: 520, loss: 0.013586846180260181
step: 530, loss: 0.082606740295887
/usr/local/lib/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=0.9073033707865168, f1=0.9010367577756833,
best_f1=0.9010367577756833
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.011628438718616962
step: 10, loss: 0.10638139396905899
step: 20, loss: 0.044413335621356964
step: 30, loss: 0.01792047545313835
step: 40, loss: 0.022399887442588806
step: 50, loss: 0.017524385824799538
step: 60, loss: 0.011720895767211914
step: 70, loss: 0.04287880286574364
step: 80, loss: 0.01883440837264061
step: 90, loss: 0.03518180921673775
step: 100, loss: 0.011912989430129528
step: 110, loss: 0.021122941747307777
step: 120, loss: 0.12130428850650787
step: 130, loss: 0.035035908222198486
step: 140, loss: 0.07643639296293259
step: 150, loss: 0.00882356334477663
step: 160, loss: 0.012884639203548431
step: 170, loss: 0.006457739509642124
step: 180, loss: 0.028297068551182747
step: 190, loss: 0.01861911080777645
step: 200, loss: 0.00842305924743414
step: 210, loss: 0.019965924322605133
step: 220, loss: 0.011563226580619812
step: 230, loss: 0.21347211301326752
step: 240, loss: 0.022502748295664787
step: 250, loss: 0.005232692696154118
step: 260, loss: 0.12625281512737274
step: 270, loss: 0.026927554979920387
step: 280, loss: 0.019112899899482727
step: 290, loss: 0.0012238946510478854
```

```
step: 300, loss: 0.005246427375823259
step: 310, loss: 0.005532054230570793
step: 320, loss: 0.029842520132660866
step: 330, loss: 0.051382407546043396
step: 340, loss: 0.037573765963315964
step: 350, loss: 0.006191709078848362
step: 360, loss: 0.01434493251144886
step: 370, loss: 0.06589564681053162
step: 380, loss: 0.012893673963844776
step: 390, loss: 0.058237772434949875
step: 400, loss: 0.04903750121593475
step: 410, loss: 0.013209155760705471
step: 420, loss: 0.007900700904428959
step: 430, loss: 0.01730825938284397
step: 440, loss: 0.020664187148213387
step: 450, loss: 0.0034342422150075436
step: 460, loss: 0.001728206523694098
step: 470, loss: 0.02082890085875988
step: 480, loss: 0.034819457679986954
step: 490, loss: 0.02694679982960224
step: 500, loss: 0.008660035207867622
step: 510, loss: 0.05405480042099953
step: 520, loss: 0.09512132406234741
step: 530, loss: 0.09151092171669006
/usr/local/lib/python3.7/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=0.8991799324650265, f1=0.8911465892597968,
best_f1=0.9010367577756833
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005539875943213701
step: 10, loss: 0.022173862904310226
step: 20, loss: 0.1690671145915985
step: 30, loss: 0.0042058550752699375
step: 40, loss: 0.04055263102054596
step: 50, loss: 0.00914285983890295
step: 60, loss: 0.18515107035636902
step: 70, loss: 0.002050961134955287
step: 80, loss: 0.00172512698918581
step: 90, loss: 0.013793508522212505
step: 100, loss: 0.006998343393206596
step: 110, loss: 0.007260981015861034
step: 120, loss: 0.06103251129388809
step: 130, loss: 0.002628639107570052
step: 140, loss: 0.0074408939108252525
step: 150, loss: 0.04326586425304413
```

```
step: 160, loss: 0.13466677069664001
step: 170, loss: 0.11165191978216171
step: 180, loss: 0.0026053208857774734
step: 190, loss: 0.005209994036704302
step: 200, loss: 0.0019380347803235054
step: 210, loss: 0.0007005877560004592
step: 220, loss: 0.0039573670364916325
step: 230, loss: 0.0032215167302638292
step: 240, loss: 0.003262768965214491
step: 250, loss: 0.0009560944745317101
step: 260, loss: 0.025056127458810806
step: 270, loss: 0.019233018159866333
step: 280, loss: 0.1498948633670807
step: 290, loss: 0.022420862689614296
step: 300, loss: 0.02787918969988823
step: 310, loss: 0.0037839370779693127
step: 320, loss: 0.02957027778029442
step: 330, loss: 0.0024238829500973225
step: 340, loss: 0.00335117825306952
step: 350, loss: 0.0009954712586477399
step: 360, loss: 0.003849540837109089
step: 370, loss: 0.001968510914593935
step: 380, loss: 0.04622243344783783
step: 390, loss: 0.01678197830915451
step: 400, loss: 0.10404867678880692
step: 410, loss: 0.005201262421905994
step: 420, loss: 0.003237547818571329
step: 430, loss: 0.01445006113499403
step: 440, loss: 0.020419608801603317
step: 450, loss: 0.005057768430560827
step: 460, loss: 0.0006478271679952741
step: 470, loss: 0.0033413933124393225
step: 480, loss: 0.00379933905787766
step: 490, loss: 0.13944652676582336
step: 500, loss: 0.013514005579054356
step: 510, loss: 0.015572095289826393
step: 520, loss: 0.018520109355449677
step: 530, loss: 0.0054131061770021915
/usr/local/lib/python3.7/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=0.8994857410004675, f1=0.8992030004688233,
best_f1=0.9010367577756833
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003760178340598941
step: 10, loss: 0.07093912363052368
```

```
step: 20, loss: 0.0055472212843596935
step: 30, loss: 0.003724275855347514
step: 40, loss: 0.0028137778863310814
step: 50, loss: 0.033647798001766205
step: 60, loss: 0.004030571319162846
step: 70, loss: 0.016154294833540916
step: 80, loss: 0.00020201240840833634
step: 90, loss: 0.0016013724962249398
step: 100, loss: 0.19362416863441467
step: 110, loss: 0.004107970278710127
step: 120, loss: 0.002116478979587555
step: 130, loss: 0.0011726325610652566
step: 140, loss: 0.00468513322994113
step: 150, loss: 0.0005880589596927166
step: 160, loss: 0.00017841279623098671
step: 170, loss: 0.0001669116463745013
step: 180, loss: 0.000192703097127378
step: 190, loss: 0.010101656429469585
step: 200, loss: 0.0011715731816366315
step: 210, loss: 0.0031821439042687416
step: 220, loss: 0.00035714704426936805
step: 230, loss: 0.011005506850779057
step: 240, loss: 0.011394361965358257
step: 250, loss: 0.00026216544210910797
step: 260, loss: 0.00046724374988116324
step: 270, loss: 0.04396532103419304
step: 280, loss: 0.03909163177013397
step: 290, loss: 0.00042764184763655066
step: 300, loss: 0.0012257697526365519
step: 310, loss: 0.0005539812264032662
step: 320, loss: 0.0011476275976747274
step: 330, loss: 0.0196986123919487
step: 340, loss: 0.008157526142895222
step: 350, loss: 0.008681879378855228
step: 360, loss: 0.00024985865456983447
step: 370, loss: 0.08918130397796631
step: 380, loss: 0.0019800690934062004
step: 390, loss: 0.0028474219143390656
step: 400, loss: 0.02593347802758217
step: 410, loss: 0.000384768150979653
step: 420, loss: 0.009618713520467281
step: 430, loss: 0.002272305777296424
step: 440, loss: 0.028569793328642845
step: 450, loss: 0.06199350580573082
step: 460, loss: 0.05499229580163956
step: 470, loss: 0.13290993869304657
step: 480, loss: 0.036652542650699615
step: 490, loss: 0.008490282110869884
```

```
step: 500, loss: 0.007875663228332996
step: 510, loss: 0.0008308067917823792
step: 520, loss: 0.0014489352470263839
step: 530, loss: 0.0011585550382733345
/usr/local/lib/python3.7/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=0.8970380818053596, f1=0.8843794242567249,
best_f1=0.9010367577756833
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1006716638803482
step: 10, loss: 0.010061459615826607
step: 20, loss: 0.007157766260206699
step: 30, loss: 0.0009074903791770339
step: 40, loss: 0.0005215282435528934
step: 50, loss: 0.00022707045718561858
step: 60, loss: 0.00027750778826884925
step: 70, loss: 0.000662682403344661
step: 80, loss: 0.00029789027757942677
step: 90, loss: 0.00019847284420393407
step: 100, loss: 0.000428570929216221
step: 110, loss: 0.03943464905023575
step: 120, loss: 0.00013005916844122112
step: 130, loss: 7.107070268830284e-05
step: 140, loss: 0.0004322124586906284
step: 150, loss: 7.925921090645716e-05
step: 160, loss: 0.034122709184885025
step: 170, loss: 0.005427144002169371
step: 180, loss: 0.00013066317478660494
step: 190, loss: 0.00321800634264946
step: 200, loss: 0.10968521982431412
step: 210, loss: 0.005138372536748648
step: 220, loss: 0.0005685415817424655
step: 230, loss: 0.0039525991305708885
step: 240, loss: 0.0009230350842699409
step: 250, loss: 0.002790557686239481
step: 260, loss: 0.004598698578774929
step: 270, loss: 6.891605880809948e-05
step: 280, loss: 0.08839989453554153
step: 290, loss: 0.0014722573105245829
step: 300, loss: 0.0002778730995487422
step: 310, loss: 0.22262823581695557
step: 320, loss: 0.011798267252743244
step: 330, loss: 0.00023015090846456587
step: 340, loss: 0.0012194806477054954
step: 350, loss: 0.0003592855064198375
```

```
step: 360, loss: 0.033826664090156555
step: 370, loss: 0.03724155202507973
step: 380, loss: 0.032383184880018234
step: 390, loss: 0.0006102664628997445
step: 400, loss: 0.0035002732183784246
step: 410, loss: 0.0025341941509395838
step: 420, loss: 0.03305184468626976
step: 430, loss: 0.005902708508074284
step: 440, loss: 8.804920798866078e-05
step: 450, loss: 0.0020447031129151583
step: 460, loss: 0.00047243168228305876
step: 470, loss: 0.14246022701263428
step: 480, loss: 0.03127482905983925
step: 490, loss: 0.0010561299277469516
step: 500, loss: 0.00023420368961524218
step: 510, loss: 0.001955082407221198
step: 520, loss: 0.004022487439215183
step: 530, loss: 0.09760626405477524
/usr/local/lib/python3.7/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=0.902123730378578, f1=0.8929236499068901,
best_f1=0.9010367577756833
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00026262958999723196
step: 10, loss: 0.012068478390574455
step: 20, loss: 0.004422382917255163
step: 30, loss: 0.006388548761606216
step: 40, loss: 0.01508947554975748
step: 50, loss: 0.00017649552319198847
step: 60, loss: 0.002825257834047079
step: 70, loss: 7.060931238811463e-05
step: 80, loss: 0.0042481557466089725
step: 90, loss: 0.0002301565691595897
step: 100, loss: 0.0002271440316690132
step: 110, loss: 0.001268636086024344
step: 120, loss: 0.00025794963585212827
step: 130, loss: 0.00015057479322422296
step: 140, loss: 4.8646179493516684e-05
step: 150, loss: 0.0011820174986496568
step: 160, loss: 0.00016223214333876967
step: 170, loss: 0.1665373295545578
step: 180, loss: 0.018078595399856567
step: 190, loss: 0.005411139689385891
step: 200, loss: 0.00015102043107617646
step: 210, loss: 0.16743136942386627
```

```
step: 220, loss: 0.002215869491919875
step: 230, loss: 0.00016640464309602976
step: 240, loss: 0.003387731732800603
step: 250, loss: 0.0006330079631879926
step: 260, loss: 0.0006014241371303797
step: 270, loss: 0.00026827698457054794
step: 280, loss: 0.0008699047612026334
step: 290, loss: 0.019417110830545425
step: 300, loss: 0.0001480371574871242
step: 310, loss: 0.053979966789484024
step: 320, loss: 0.019578304141759872
step: 330, loss: 0.028791315853595734
step: 340, loss: 0.0007202147389762104
step: 350, loss: 0.0006071575917303562
step: 360, loss: 0.00047164823627099395
step: 370, loss: 0.009273711591959
step: 380, loss: 0.11564601212739944
step: 390, loss: 0.0011114060180261731
step: 400, loss: 0.005951728206127882
step: 410, loss: 0.09336038678884506
step: 420, loss: 0.0005846221465617418
step: 430, loss: 0.0002512976061552763
step: 440, loss: 0.006468301173299551
step: 450, loss: 0.00019276388047728688
step: 460, loss: 0.001595796667970717
step: 470, loss: 0.000565899710636586
step: 480, loss: 0.004885638598352671
step: 490, loss: 0.00023944920394569635
step: 500, loss: 0.001492420444265008
step: 510, loss: 0.01848262920975685
step: 520, loss: 0.0015927230706438422
step: 530, loss: 0.00035365470102988183
/usr/local/lib/python3.7/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=0.8991954566966399, f1=0.8840996168582376,
best_f1=0.9010367577756833
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002653761301189661
step: 10, loss: 0.001569288084283471
step: 20, loss: 0.0003667429555207491
step: 30, loss: 0.012460277415812016
step: 40, loss: 0.0003895804693456739
step: 50, loss: 0.0015609057154506445
step: 60, loss: 0.00013434034190140665
step: 70, loss: 0.08426154404878616
```

```
step: 80, loss: 0.0005193522665649652
step: 90, loss: 0.0042283437214791775
step: 100, loss: 0.0013826247304677963
step: 110, loss: 0.0003680252411868423
step: 120, loss: 0.0002496697998140007
step: 130, loss: 8.947340393206105e-05
step: 140, loss: 0.0009827063186094165
step: 150, loss: 0.001690963632427156
step: 160, loss: 0.00011251957766944543
step: 170, loss: 0.0011489324970170856
step: 180, loss: 5.307763422024436e-05
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 190, loss: 0.14743882417678833
step: 200, loss: 0.00043059090967290103
step: 210, loss: 0.0002168121573049575
step: 220, loss: 0.0020965179428458214
step: 230, loss: 0.00021255257888697088
step: 240, loss: 0.004001119174063206
step: 250, loss: 0.0580708347260952
step: 260, loss: 0.0009948932565748692
step: 270, loss: 0.004432446323335171
step: 280, loss: 0.0006297242944128811
step: 290, loss: 0.00011728171375580132
step: 300, loss: 0.0052511501125991344
step: 310, loss: 0.00011987549805780873
step: 320, loss: 0.00022356075351126492
step: 330, loss: 0.0011144927702844143
step: 340, loss: 7.06135033397004e-05
step: 350, loss: 5.898510789847933e-05
step: 360, loss: 0.008208216167986393
step: 370, loss: 0.0691952034831047
step: 380, loss: 0.00012753524060826749
step: 390, loss: 0.00017599428247194737
step: 400, loss: 0.042549166828393936
step: 410, loss: 0.000544345413800329
step: 420, loss: 0.0006795627996325493
step: 430, loss: 0.00012832321226596832
step: 440, loss: 0.0001987520226975903
step: 450, loss: 0.0017523982096463442
step: 460, loss: 0.02053011767566204
step: 470, loss: 0.005663459189236164
step: 480, loss: 0.00011090083717135713
step: 490, loss: 0.0009433956583961844
step: 500, loss: 0.0017456711502745748
step: 510, loss: 0.00022256645024754107
step: 520, loss: 6.069739174563438e-05
step: 530, loss: 0.0006614120211452246
/usr/local/lib/python3.7/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=0.89900426742532, f1=0.8883587786259542,
best_f1=0.9010367577756833
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00018567751976661384
step: 10, loss: 2.8035996365360916e-05
step: 20, loss: 0.0005400180816650391
step: 30, loss: 5.288731335895136e-05
step: 40, loss: 0.009004865773022175
step: 50, loss: 0.0001903182128444314
step: 60, loss: 5.389509169617668e-05
step: 70, loss: 9.140982729149982e-05
step: 80, loss: 0.007065861020237207
step: 90, loss: 8.580060239182785e-05
step: 100, loss: 0.00734102725982666
step: 110, loss: 0.0020992346107959747
step: 120, loss: 0.001569963525980711
step: 130, loss: 0.002616242738440633
step: 140, loss: 7.042875949991867e-05
step: 150, loss: 0.00011666089994832873
step: 160, loss: 0.00010938890045508742
step: 170, loss: 0.0031670306343585253
step: 180, loss: 0.15206220746040344
step: 190, loss: 0.0004234028165228665
step: 200, loss: 0.0030468706972897053
step: 210, loss: 0.006380421109497547
step: 220, loss: 0.00018839236872736365
step: 230, loss: 0.0012321770191192627
step: 240, loss: 0.00016350667283404619
step: 250, loss: 0.0007382985786534846
step: 260, loss: 0.026797188445925713
step: 270, loss: 0.000872669043019414
step: 280, loss: 0.00017228606157004833
step: 290, loss: 0.0029061275999993086
step: 300, loss: 0.006021385081112385
step: 310, loss: 0.0003723542613442987
step: 320, loss: 0.07731960713863373
step: 330, loss: 0.0003460282750893384
step: 340, loss: 0.00020218550343997777
step: 350, loss: 0.0009674594621174037
step: 360, loss: 0.00022483702923636883
step: 370, loss: 0.0005131378420628607
step: 380, loss: 0.00041432440048083663
step: 390, loss: 0.0021063745953142643
step: 400, loss: 0.0020649598445743322
```

```
step: 410, loss: 0.00039692362770438194
step: 420, loss: 5.248490924714133e-05
step: 430, loss: 6.43537932774052e-05
step: 440, loss: 3.861896402668208e-05
step: 450, loss: 0.00019866494403686374
step: 460, loss: 8.97282880032435e-05
step: 470, loss: 0.00022195407655090094
step: 480, loss: 8.157156116794795e-05
step: 490, loss: 0.10406329482793808
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 500, loss: 0.00011815728794317693
step: 510, loss: 0.0005042797420173883
step: 520, loss: 0.00014088614261709154
step: 530, loss: 0.00036622327752411366
/usr/local/lib/python3.7/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=0.9011095031355524, f1=0.8729609490855166,
best_f1=0.9010367577756833
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00012673699529841542
step: 10, loss: 0.0033488881308585405
step: 20, loss: 0.0011087703751400113
step: 30, loss: 0.00036186579382047057
step: 40, loss: 5.228550435276702e-05
step: 50, loss: 0.00010642228153301403
step: 60, loss: 3.529537207214162e-05
step: 70, loss: 7.862656639190391e-05
step: 80, loss: 5.494219658430666e-05
step: 90, loss: 6.2868173699826e-05
step: 100, loss: 3.228236164432019e-05
step: 110, loss: 5.3674910304835066e-05
step: 120, loss: 6.952081457711756e-05
step: 130, loss: 5.225716085988097e-05
step: 140, loss: 4.3683718104148284e-05
step: 150, loss: 3.4115251764887944e-05
step: 160, loss: 8.072747732512653e-05
step: 170, loss: 0.00013810979726258665
step: 180, loss: 2.310019772266969e-05
step: 190, loss: 9.502654575044289e-05
step: 200, loss: 0.0004389439127407968
step: 210, loss: 3.5675624530995265e-05
step: 220, loss: 0.000298016908345744
step: 230, loss: 3.8104735722299665e-05
step: 240, loss: 2.7778931325883605e-05
step: 250, loss: 0.0005761768552474678
```

```
step: 260, loss: 3.4498374589020386e-05
step: 270, loss: 0.00043141868081875145
step: 280, loss: 0.002853243611752987
step: 290, loss: 0.00010713120718719438
step: 300, loss: 0.00013891990238334984
step: 310, loss: 0.015654245391488075
step: 320, loss: 0.0006780445692129433
step: 330, loss: 0.0019025546498596668
step: 340, loss: 2.4381612092838623e-05
step: 350, loss: 0.0005999594577588141
step: 360, loss: 0.0003216999175492674
step: 370, loss: 0.00023597570543643087
step: 380, loss: 0.006831992417573929
step: 390, loss: 0.001490720547735691
step: 400, loss: 0.0014098346000537276
step: 410, loss: 2.2567490304936655e-05
step: 420, loss: 0.00463460385799408
step: 430, loss: 3.185334207955748e-05
step: 440, loss: 0.00013785959163215011
step: 450, loss: 2.5305482267867774e-05
step: 460, loss: 0.09656257182359695
step: 470, loss: 0.0006012713420204818
step: 480, loss: 0.0002561819856055081
step: 490, loss: 0.002083658240735531
step: 500, loss: 6.561349437106401e-05
step: 510, loss: 5.8566234656609595e-05
step: 520, loss: 3.0706371035194024e-05
step: 530, loss: 3.1659830710850656e-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=0.8952830188679246, f1=0.8875, best_f1=0.9010367577756833
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009968921076506376
step: 10, loss: 7.32696134946309e-05
step: 20, loss: 2.8598418793990277e-05
step: 30, loss: 3.994730286649428e-05
step: 40, loss: 0.0015480684814974666
step: 50, loss: 0.0009748090524226427
step: 60, loss: 6.614008452743292e-05
step: 70, loss: 0.00011210133379790932
step: 80, loss: 0.00024656925234012306
step: 90, loss: 0.00012889050412923098
step: 100, loss: 0.00022410383098758757
step: 110, loss: 2.4109647711156867e-05
step: 120, loss: 0.0008389588329009712
```

```
step: 130, loss: 3.503787957015447e-05
step: 140, loss: 5.722701098420657e-05
step: 150, loss: 0.0021576841827481985
step: 160, loss: 3.697485590237193e-05
step: 170, loss: 4.0798975533107296e-05
step: 180, loss: 4.366682696854696e-05
step: 190, loss: 2.096556272590533e-05
step: 200, loss: 4.572779289446771e-05
step: 210, loss: 8.830869774101302e-05
step: 220, loss: 0.0012662699446082115
step: 230, loss: 0.00017365740495733917
step: 240, loss: 2.5595936676836573e-05
step: 250, loss: 0.0002679743920452893
step: 260, loss: 0.0005628978833556175
step: 270, loss: 0.00012196910392958671
step: 280, loss: 3.823516453849152e-05
step: 290, loss: 0.0008354483870789409
step: 300, loss: 0.010734152048826218
step: 310, loss: 6.128466338850558e-05
step: 320, loss: 0.00027894743834622204
step: 330, loss: 6.183759978739545e-05
step: 340, loss: 7.130236917873845e-05
step: 350, loss: 0.0003576385206542909
step: 360, loss: 0.005340684205293655
step: 370, loss: 5.192149546928704e-05
step: 380, loss: 0.00027348092407919466
step: 390, loss: 4.0986622479977086e-05
step: 400, loss: 3.759734318009578e-05
step: 410, loss: 0.0003543886705301702
step: 420, loss: 5.4805288527859375e-05
step: 430, loss: 3.823880251729861e-05
step: 440, loss: 9.34850177145563e-05
step: 450, loss: 0.00033825053833425045
step: 460, loss: 0.00024482241133227944
step: 470, loss: 0.0005236067227087915
step: 480, loss: 0.00015816812810953707
step: 490, loss: 0.0030529522337019444
step: 500, loss: 0.0001667961332714185
step: 510, loss: 4.81095521536190006e-05
step: 520, loss: 0.015967490151524544
step: 530, loss: 6.780654075555503e-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=0.9003306565895135, f1=0.888888888888889,
best_f1=0.9010367577756833
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.5347791203530505e-05
step: 10, loss: 3.809457120951265e-05
step: 20, loss: 0.0002583693712949753
step: 30, loss: 0.0022858199663460255
step: 40, loss: 6.154160655569285e-05
step: 50, loss: 3.633486630860716e-05
step: 60, loss: 7.445477967849001e-05
step: 70, loss: 9.712432802189142e-05
step: 80, loss: 3.590649794205092e-05
step: 90, loss: 5.9558475186349824e-05
step: 100, loss: 0.0001733774261083454
step: 110, loss: 5.210713061387651e-05
step: 120, loss: 5.0289661885472015e-05
step: 130, loss: 6.066774221835658e-05
step: 140, loss: 3.171236676280387e-05
step: 150, loss: 4.1054398025153205e-05
step: 160, loss: 0.00011460013047326356
step: 170, loss: 0.00012346560833975673
step: 180, loss: 0.00013212299381848425
step: 190, loss: 8.871047612046823e-05
step: 200, loss: 0.002077449345961213
step: 210, loss: 0.001808093162253499
step: 220, loss: 3.367129829712212e-05
step: 230, loss: 0.0007155759958550334
step: 240, loss: 3.298959927633405e-05
step: 250, loss: 0.009622576646506786
step: 260, loss: 0.00013461499474942684
step: 270, loss: 4.568340227706358e-05
step: 280, loss: 0.00012348112068139017
step: 290, loss: 0.013424372300505638
step: 300, loss: 6.690374721074477e-05
step: 310, loss: 2.19527464651037e-05
step: 320, loss: 2.1528076104004867e-05
step: 330, loss: 4.147016443312168e-05
step: 340, loss: 2.6866049665841274e-05
step: 350, loss: 0.0003099808527622372
step: 360, loss: 0.000140607007779181
step: 370, loss: 0.0002892881166189909
step: 380, loss: 0.0012452988885343075
step: 390, loss: 2.2220949176698923e-05
step: 400, loss: 2.1606307200272568e-05
step: 410, loss: 4.4427939428715035e-05
step: 420, loss: 9.250792209059e-05
step: 430, loss: 3.348102472955361e-05
step: 440, loss: 0.0037369902711361647
step: 450, loss: 3.9214010030264035e-05
step: 460, loss: 2.0086481526959687e-05
```

```
step: 470, loss: 3.420401117182337e-05
step: 480, loss: 3.431528602959588e-05
step: 490, loss: 0.00014841670054011047
step: 500, loss: 0.007499702274799347
step: 510, loss: 2.5837951397988945e-05
step: 520, loss: 0.00023273189435712993
step: 530, loss: 3.032665881619323e-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=0.902671755725191, f1=0.8875, best_f1=0.9010367577756833
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005899170413613319
step: 10, loss: 4.769543738802895e-05
step: 20, loss: 8.7280037405435e-05
step: 30, loss: 8.046761649893597e-05
step: 40, loss: 9.267444693250582e-05
step: 50, loss: 4.5954202505527064e-05
step: 60, loss: 4.707471816800535e-05
step: 70, loss: 2.5022112822625786e-05
step: 80, loss: 0.00013065009261481464
step: 90, loss: 7.640883995918557e-05
step: 100, loss: 8.789185812929645e-05
step: 110, loss: 2.3785549274180084e-05
step: 120, loss: 5.5874796089483425e-05
step: 130, loss: 0.0005772595177404583
step: 140, loss: 4.0268245356855914e-05
step: 150, loss: 5.3760071750730276e-05
step: 160, loss: 0.0003186299873050302
step: 170, loss: 0.0006922595202922821
step: 180, loss: 3.424768874538131e-05
step: 190, loss: 0.00011263177293585613
step: 200, loss: 3.1466042855754495e-05
step: 210, loss: 7.90377234807238e-05
step: 220, loss: 0.00017776561435312033
step: 230, loss: 3.945163552998565e-05
step: 240, loss: 8.515120862284675e-05
step: 250, loss: 2.2530159185407683e-05
step: 260, loss: 2.9452765375026502e-05
step: 270, loss: 0.0017891849856823683
step: 280, loss: 3.646792174549773e-05
step: 290, loss: 0.00011237911530770361
step: 300, loss: 0.001688729040324688
step: 310, loss: 3.18156489811372e-05
step: 320, loss: 0.00037942989729344845
step: 330, loss: 5.759525447501801e-05
```

```
step: 340, loss: 3.5631994251161814e-05
step: 350, loss: 0.00010051445860881358
step: 360, loss: 0.0005257803131826222
step: 370, loss: 2.793866951833479e-05
step: 380, loss: 2.8202153771417215e-05
step: 390, loss: 2.710395710892044e-05
step: 400, loss: 0.00014869397273287177
step: 410, loss: 0.00015048269415274262
step: 420, loss: 1.7948261302080937e-05
step: 430, loss: 2.818100074364338e-05
step: 440, loss: 2.2808866560808383e-05
step: 450, loss: 2.4388953534071334e-05
step: 460, loss: 5.115141539135948e-05
step: 470, loss: 5.1164402975700796e-05
step: 480, loss: 2.64822956523858e-05
step: 490, loss: 2.5010822355397977e-05
step: 500, loss: 4.0208127757068723e-05
step: 510, loss: 3.2992818887578323e-05
step: 520, loss: 3.893371831509285e-05
step: 530, loss: 3.165613452438265e-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=0.9018867924528302, f1=0.8901569186875892,
best_f1=0.9010367577756833
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 2.1803658455610275e-05
step: 10, loss: 3.4679869713727385e-05
step: 20, loss: 3.764824941754341e-05
step: 30, loss: 1.917010740726255e-05
step: 40, loss: 0.002519866917282343
step: 50, loss: 0.00019715684175025672
step: 60, loss: 1.966164018085692e-05
step: 70, loss: 3.1044677598401904e-05
step: 80, loss: 9.341266559204087e-05
step: 90, loss: 2.4373337510041893e-05
step: 100, loss: 2.6773002900881693e-05
step: 110, loss: 0.023883510380983353
step: 120, loss: 0.000476348795928061
step: 130, loss: 0.00010555365588515997
step: 140, loss: 1.5340594472945668e-05
step: 150, loss: 0.00012737087672576308
step: 160, loss: 3.977943561039865e-05
step: 170, loss: 4.336818165029399e-05
step: 180, loss: 5.442549809231423e-05
step: 190, loss: 4.56758207292296e-05
```

```
step: 200, loss: 2.9200187782407738e-05
step: 210, loss: 2.3595352104166523e-05
step: 220, loss: 1.4614192878070753e-05
step: 230, loss: 2.955742456833832e-05
step: 240, loss: 5.51086341147311e-05
step: 250, loss: 4.301282388041727e-05
step: 260, loss: 8.454315684502944e-05
step: 270, loss: 3.221826773369685e-05
step: 280, loss: 0.0007717356784269214
step: 290, loss: 2.8854577976744622e-05
step: 300, loss: 0.020425403490662575
step: 310, loss: 2.753615081019234e-05
step: 320, loss: 3.390873462194577e-05
step: 330, loss: 0.0029972316697239876
step: 340, loss: 0.0008243973716162145
step: 350, loss: 4.777988579007797e-05
step: 360, loss: 3.0877767130732536e-05
step: 370, loss: 1.7881193343782797e-05
step: 380, loss: 2.0496001525316387e-05
step: 390, loss: 2.000069071073085e-05
step: 400, loss: 0.0009195629390887916
step: 410, loss: 0.0002105372550431639
step: 420, loss: 2.0924671844113618e-05
step: 430, loss: 2.609071452752687e-05
step: 440, loss: 0.0007093353196978569
step: 450, loss: 0.00011124831507913768
step: 460, loss: 2.627649882924743e-05
step: 470, loss: 0.00012497570423875004
step: 480, loss: 2.6145695301238447e-05
step: 490, loss: 0.00010252720676362514
step: 500, loss: 6.906674389028922e-05
step: 510, loss: 0.0004617772647179663
step: 520, loss: 3.179049235768616e-05
step: 530, loss: 1.796684409782756e-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=0.9017175572519084, f1=0.8885658914728681,
best_f1=0.9010367577756833
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.")
5742it [00:14, 395.38it/s]
load_f1 = 0.9019788311090658
real_f1 = 0.902450300508553
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, 388.79it/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 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: 0.8681785464286804
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=0.3111111111111111, f1=0.2857142857142857,
best_f1=0.2857142857142857
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.37089288234710693
/usr/local/lib/python3.7/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=0.2857142857142857, f1=0.2745098039215686,
best_f1=0.2857142857142857
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3549487292766571
/usr/local/lib/python3.7/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=0.27184466019417475, f1=0.2692307692307693,
best_f1=0.2857142857142857
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.37196940183639526
/usr/local/lib/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=0.41935483870967744, f1=0.2891566265060241,
best_f1=0.2891566265060241
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.275824636220932
/usr/local/lib/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=0.4705882352941177, f1=0.34374999999999994,
best_f1=0.34374999999999994
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2657514810562134
/usr/local/lib/python3.7/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=0.4444444444444444, f1=0.41379310344827586,
best_f1=0.34374999999999994
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2273828685283661
/usr/local/lib/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=0.4897959183673469, f1=0.4067796610169491,
best_f1=0.4067796610169491
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.38199383020401
/usr/local/lib/python3.7/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=0.48888888888888893, f1=0.42857142857142855,
best_f1=0.4067796610169491
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.22274532914161682
/usr/local/lib/python3.7/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=0.4727272727272727, f1=0.4745762711864407,
best_f1=0.4067796610169491
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2553536295890808
/usr/local/lib/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=0.5217391304347825, f1=0.42857142857142855,
best_f1=0.42857142857142855
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.21962115168571472
/usr/local/lib/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=0.56, f1=0.47058823529411764, best_f1=0.47058823529411764
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2702135443687439
/usr/local/lib/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=0.5833333333333334, f1=0.4571428571428571,
```

```
best_f1=0.4571428571428571
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.14939351379871368
/usr/local/lib/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=0.64, f1=0.47058823529411764, best_f1=0.47058823529411764
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2216826230287552
/usr/local/lib/python3.7/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=0.64, f1=0.4571428571428571, best_f1=0.47058823529411764
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2302975356578827
/usr/local/lib/python3.7/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=0.64, f1=0.4571428571428571, best_f1=0.47058823529411764
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, 114105.13it/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 = 0.5294117647058824
real_f1 = 0.5142857142857143
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, 424.84it/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 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: 0.8008906841278076
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0

```
step: 10, loss: 0.49258244037628174
step: 20, loss: 0.6089513301849365
step: 30, loss: 0.5105851292610168
step: 40, loss: 0.4511651396751404
step: 50, loss: 0.3662388324737549
step: 60, loss: 0.29804471135139465
step: 70, loss: 0.14558997750282288
step: 80, loss: 0.07524562627077103
step: 90, loss: 0.009958893992006779
step: 100, loss: 0.0638984888792038
step: 110, loss: 0.010421739891171455
step: 120, loss: 0.031301841139793396
step: 130, loss: 0.17467671632766724
step: 140, loss: 0.03312826529145241
step: 150, loss: 0.04910871386528015
step: 160, loss: 0.11703725159168243
step: 170, loss: 0.021919064223766327
step: 180, loss: 0.03741803392767906
step: 190, loss: 0.3482377827167511
step: 200, loss: 0.030877824872732162
step: 210, loss: 0.026674211025238037
step: 220, loss: 0.035566989332437515
step: 230, loss: 0.037579141557216644
/usr/local/lib/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=0.9622222222222223, f1=0.961625282167043,
best_f1=0.961625282167043
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.030298730358481407
step: 10, loss: 0.0015709699364379048
step: 20, loss: 0.03306726738810539
step: 30, loss: 0.004060381092131138
step: 40, loss: 0.03387844190001488
step: 50, loss: 0.0024172738194465637
step: 60, loss: 0.12644165754318237
step: 70, loss: 0.02073793113231659
step: 80, loss: 0.009692663326859474
step: 90, loss: 0.008979114703834057
step: 100, loss: 0.03556143119931221
step: 110, loss: 0.06333428621292114
step: 120, loss: 0.07772713154554367
```

```
step: 130, loss: 0.049290306866168976
step: 140, loss: 0.015474338084459305
step: 150, loss: 0.006832794286310673
step: 160, loss: 0.01141358818858862
step: 170, loss: 0.20973743498325348
step: 180, loss: 0.009205206297338009
step: 190, loss: 0.0255043376237154
step: 200, loss: 0.004261136054992676
step: 210, loss: 0.03996874764561653
step: 220, loss: 0.0010629236930981278
step: 230, loss: 0.1865631639957428
/usr/local/lib/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=0.9723145071982282, 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: 0.07753852009773254
step: 10, loss: 0.053970541805028915
step: 20, loss: 0.008245579898357391
step: 30, loss: 0.013133893720805645
step: 40, loss: 0.010669790208339691
step: 50, loss: 0.04992109164595604
step: 60, loss: 0.02977827377617359
step: 70, loss: 0.00371891469694674
step: 80, loss: 0.015700364485383034
step: 90, loss: 0.025319620966911316
step: 100, loss: 0.004651317838579416
step: 110, loss: 0.16614404320716858
step: 120, loss: 0.021322093904018402
step: 130, loss: 0.003283932339400053
step: 140, loss: 0.010268252342939377
step: 150, loss: 0.0371820405125618
step: 160, loss: 0.004760660231113434
step: 170, loss: 0.016985571011900902
step: 180, loss: 0.007153964601457119
step: 190, loss: 0.020263442769646645
step: 200, loss: 0.003419114276766777
step: 210, loss: 0.04125913977622986
step: 220, loss: 0.00678770849481225
step: 230, loss: 0.06175587698817253
/usr/local/lib/python3.7/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=0.9689578713968958, f1=0.9689578713968958,
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: 0.0014000978553667665
step: 10, loss: 0.0037144639063626528
step: 20, loss: 0.0051808543503284454
step: 30, loss: 0.0009898346615955234
step: 40, loss: 0.004182893317192793
step: 50, loss: 0.009851709008216858
step: 60, loss: 0.005811662878841162
step: 70, loss: 0.017459426075220108
step: 80, loss: 0.04452252760529518
step: 90, loss: 0.056147608906030655
step: 100, loss: 0.02826179377734661
step: 110, loss: 0.03340616822242737
step: 120, loss: 0.010187164880335331
step: 130, loss: 0.003221136052161455
step: 140, loss: 0.0025603552348911762
step: 150, loss: 0.0007357136928476393
step: 160, loss: 0.0035150996409356594
step: 170, loss: 0.006125818006694317
step: 180, loss: 0.05456704646348953
step: 190, loss: 0.006736688315868378
step: 200, loss: 0.0023568077012896538
step: 210, loss: 0.0008499989053234458
step: 220, loss: 0.0014322769129648805
step: 230, loss: 0.04239365831017494
/usr/local/lib/python3.7/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=0.9710467706013363, f1=0.9711751662971175,
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: 0.002473984844982624
step: 10, loss: 0.00519841443747282
step: 20, loss: 0.03335786238312721
step: 30, loss: 0.07401362806558609
step: 40, loss: 0.013196263462305069
step: 50, loss: 0.0006622842629440129
step: 60, loss: 0.00016290748317260295
step: 70, loss: 0.00037193234311416745
step: 80, loss: 0.06261610239744186
```

```
step: 90, loss: 0.023884519934654236
step: 100, loss: 0.0100499102845788
step: 110, loss: 0.00852509681135416
step: 120, loss: 0.06917686760425568
step: 130, loss: 0.0338449664413929
step: 140, loss: 0.0005391016020439565
step: 150, loss: 0.00042534081148914993
step: 160, loss: 0.0167087409645319
step: 170, loss: 0.12049739062786102
step: 180, loss: 0.0042540510185062885
step: 190, loss: 0.0005123170558363199
step: 200, loss: 0.019625939428806305
step: 210, loss: 0.003254529321566224
step: 220, loss: 0.004446927923709154
step: 230, loss: 0.0005432270118035376
/usr/local/lib/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=0.9742441209406495, f1=0.9683972911963882,
best_f1=0.9683972911963882
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008944649016484618
step: 10, loss: 0.001127816503867507
step: 20, loss: 0.0005514916847459972
step: 30, loss: 0.0004278359701856971
step: 40, loss: 0.00772458640858531
step: 50, loss: 0.00109999134974554181
step: 60, loss: 0.0177697092294693
step: 70, loss: 0.0018083018949255347
step: 80, loss: 0.00491287000477314
step: 90, loss: 0.00033778860233724117
step: 100, loss: 0.00010324896720703691
step: 110, loss: 0.07444257289171219
step: 120, loss: 0.0006485146004706621
step: 130, loss: 0.0004471586726140231
step: 140, loss: 0.016990557312965393
step: 150, loss: 0.0031235527712851763
step: 160, loss: 0.01333669014275074
step: 170, loss: 0.0005931280320510268
step: 180, loss: 0.0013928498374298215
step: 190, loss: 0.0064828782342374325
step: 200, loss: 0.02022126317024231
```

```
step: 210, loss: 0.0003319647803436965
step: 220, loss: 0.0004084069805685431
step: 230, loss: 0.00016558753850404173
/usr/local/lib/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=0.9797752808988766, f1=0.967670011148272,
best_f1=0.967670011148272
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.044319331645965576
step: 10, loss: 0.00021371273032855242
step: 20, loss: 0.0005053287604823709
step: 30, loss: 0.000874430756084621
step: 40, loss: 0.000155607151100412
step: 50, loss: 9.087514627026394e-05
step: 60, loss: 0.004642682615667586
step: 70, loss: 0.016280358657240868
step: 80, loss: 0.009601768106222153
step: 90, loss: 0.0010960021754726768
step: 100, loss: 0.008327542804181576
step: 110, loss: 0.012646033428609371
step: 120, loss: 0.0002888813032768667
step: 130, loss: 0.0005487108719535172
step: 140, loss: 0.0005868448060937226
step: 150, loss: 0.0016238511307165027
step: 160, loss: 0.00023303418129216880
step: 170, loss: 0.00016079883789643645
step: 180, loss: 0.00924212671816349
step: 190, loss: 0.021623268723487854
step: 200, loss: 0.01174146868288517
step: 210, loss: 0.00013605302956420928
step: 220, loss: 0.000783480703830719
step: 230, loss: 0.024310756474733353
/usr/local/lib/python3.7/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=0.9698996655518396, f1=0.9665924276169264,
best_f1=0.967670011148272
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0325370617210865
```

```
step: 10, loss: 0.03208096697926521
step: 20, loss: 0.00018091057427227497
step: 30, loss: 0.00012456387048587203
step: 40, loss: 0.0004758075810968876
step: 50, loss: 0.0002413627371424809
step: 60, loss: 7.421285408781841e-05
step: 70, loss: 0.00019996306218672544
step: 80, loss: 0.0015162023482844234
step: 90, loss: 0.0008046245202422142
step: 100, loss: 7.202091364888474e-05
step: 110, loss: 0.006502984557300806
step: 120, loss: 0.000335335178533569
step: 130, loss: 0.01622764579951763
step: 140, loss: 8.844503463478759e-05
step: 150, loss: 0.0624002106487751
step: 160, loss: 0.0008382551604881883
step: 170, loss: 9.632504952605814e-05
step: 180, loss: 0.013408687897026539
step: 190, loss: 0.0016336950939148664
step: 200, loss: 0.001495201955549419
step: 210, loss: 7.986713171703741e-05
step: 220, loss: 0.0022520185448229313
step: 230, loss: 0.023109711706638336
/usr/local/lib/python3.7/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=0.9785794813979707, f1=0.9665924276169264,
best_f1=0.967670011148272
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0026096946094185114
step: 10, loss: 8.779771451372653e-05
step: 20, loss: 0.0003411069919820875
step: 30, loss: 0.0005756299360655248
step: 40, loss: 0.003540599951520562
step: 50, loss: 0.00013592773757409304
step: 60, loss: 0.00011205569899175316
step: 70, loss: 0.0011088326573371887
step: 80, loss: 0.04752138629555702
step: 90, loss: 0.007669596932828426
step: 100, loss: 0.0001122577377827838
step: 110, loss: 0.0007560760132037103
step: 120, loss: 0.013500125147402287
step: 130, loss: 0.00010399722668807954
step: 140, loss: 0.003146954346448183
step: 150, loss: 7.455705053871498e-05
step: 160, loss: 0.004595446866005659
```

```
step: 170, loss: 0.000818534754216671
step: 180, loss: 0.010446294210851192
step: 190, loss: 6.017863779561594e-05
step: 200, loss: 0.0005618135910481215
step: 210, loss: 0.0046147918328642845
step: 220, loss: 0.025565307587385178
step: 230, loss: 0.00546693429350853
/usr/local/lib/python3.7/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=0.9753914988814317, f1=0.9656699889258028,
best_f1=0.967670011148272
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0021256848704069853
step: 10, loss: 8.5852800111752e-05
step: 20, loss: 0.0478050522506237
step: 30, loss: 0.001396358828060329
step: 40, loss: 0.006119458936154842
step: 50, loss: 0.015329397283494473
step: 60, loss: 0.00049482064787729844
step: 70, loss: 0.0001482867228332907
step: 80, loss: 0.023825079202651978
step: 90, loss: 0.004053385928273201
step: 100, loss: 0.0002089149784296751
step: 110, loss: 0.00029015069594606757
step: 120, loss: 0.012830853462219238
step: 130, loss: 0.002182260388508439
step: 140, loss: 0.005266754422336817
step: 150, loss: 0.0013516958570107818
step: 160, loss: 0.00038342594052664936
step: 170, loss: 0.001617770059965551
step: 180, loss: 0.0006917246500961483
step: 190, loss: 0.000144786637974903
step: 200, loss: 0.00046497059520334005
step: 210, loss: 0.0002981375146191567
step: 220, loss: 0.05180314928293228
step: 230, loss: 6.0506994486786425e-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=0.9741863075196409, f1=0.9698996655518396,
best_f1=0.967670011148272
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.691362068522722e-05
```

```
step: 10, loss: 0.00032189980265684426
step: 20, loss: 0.00015757459914311767
step: 30, loss: 0.0027918301057070494
step: 40, loss: 0.0029398060869425535
step: 50, loss: 0.0007259403355419636
step: 60, loss: 0.00438049016520381
step: 70, loss: 0.12964454293251038
step: 80, loss: 0.00018410268239676952
step: 90, loss: 0.00011987437756033614
step: 100, loss: 0.00020658536232076585
step: 110, loss: 4.348342554294504e-05
step: 120, loss: 0.0014444557018578053
step: 130, loss: 0.0002599285217002034
step: 140, loss: 6.975572614464909e-05
step: 150, loss: 5.2483341278275475e-05
step: 160, loss: 8.899635577108711e-05
step: 170, loss: 0.004372553899884224
step: 180, loss: 0.002142789773643017
step: 190, loss: 0.010551055893301964
step: 200, loss: 0.0011920193210244179
step: 210, loss: 4.695697862189263e-05
step: 220, loss: 0.00019824603805318475
step: 230, loss: 0.02063286490738392
/usr/local/lib/python3.7/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=0.9753914988814317, f1=0.96875, best_f1=0.967670011148272
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0011547288158908486
step: 10, loss: 0.0001322521857218817
step: 20, loss: 5.4366046242648736e-05
step: 30, loss: 0.0005177740240469575
step: 40, loss: 5.953924483037554e-05
step: 50, loss: 9.253650932805613e-05
step: 60, loss: 0.001286118756979704
step: 70, loss: 0.0018621193012222648
step: 80, loss: 9.472314559388906e-05
step: 90, loss: 3.0338071155711077e-05
step: 100, loss: 0.00025342925800941885
step: 110, loss: 0.0055036162957549095
step: 120, loss: 0.0002806222182698548
step: 130, loss: 0.00044442559010349214
step: 140, loss: 5.6387565564364195e-05
step: 150, loss: 0.00011190425720997155
step: 160, loss: 0.02249736152589321
step: 170, loss: 4.715430623036809e-05
```

```
step: 180, loss: 3.877437120536342e-05
step: 190, loss: 0.0013211960904300213
step: 200, loss: 0.004711721558123827
step: 210, loss: 2.946645145129878e-05
step: 220, loss: 7.755469414405525e-05
step: 230, loss: 3.710640157805756e-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=0.9740698985343857, f1=0.970917225950783,
best_f1=0.967670011148272
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0020273257978260517
step: 10, loss: 7.034639565972611e-05
step: 20, loss: 4.300966247683391e-05
step: 30, loss: 0.008791731670498848
step: 40, loss: 6.442174344556406e-05
step: 50, loss: 3.414526509004645e-05
step: 60, loss: 0.0001264770544366911
step: 70, loss: 3.501311221043579e-05
step: 80, loss: 2.9187069230829366e-05
step: 90, loss: 5.877812873222865e-05
step: 100, loss: 0.00011764825467253104
step: 110, loss: 7.411031401716173e-05
step: 120, loss: 8.998569683171809e-05
step: 130, loss: 0.00020616903202608228
step: 140, loss: 0.000739079259801656
step: 150, loss: 0.004006121773272753
step: 160, loss: 3.9090733480406925e-05
step: 170, loss: 4.576524588628672e-05
step: 180, loss: 0.0002242780610686168
step: 190, loss: 3.97396543121431e-05
step: 200, loss: 0.00012532496475614607
step: 210, loss: 0.000236681240494363
step: 220, loss: 0.0002056544180959463
step: 230, loss: 0.000357151962816671524
/usr/local/lib/python3.7/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=0.9743589743589743, f1=0.9700996677740864,
best_f1=0.967670011148272
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.837024845182896e-05
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
```

```
step: 10, loss: 0.00030826020520180464
step: 20, loss: 0.0005661172908730805
step: 30, loss: 4.48876635346096e-05
step: 40, loss: 2.609147850307636e-05
step: 50, loss: 9.728903387440369e-05
step: 60, loss: 6.583380309166387e-05
step: 70, loss: 0.004341835156083107
step: 80, loss: 2.8650552849285305e-05
step: 90, loss: 8.699100726516917e-05
step: 100, loss: 0.0019140300573781133
step: 110, loss: 6.821937131462619e-05
step: 120, loss: 0.001143977278843522
step: 130, loss: 8.355638419743627e-05
step: 140, loss: 0.00030842673731967807
step: 150, loss: 3.907270365743898e-05
step: 160, loss: 9.637501352699474e-05
step: 170, loss: 4.94400410389062e-05
step: 180, loss: 0.00012477970449253917
step: 190, loss: 0.0002850544115062803
step: 200, loss: 5.03504961670842e-05
step: 210, loss: 0.0009960309835150838
step: 220, loss: 7.248005567817017e-05
step: 230, loss: 3.4638513170648366e-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=0.9732739420935412, f1=0.9711751662971175,
best_f1=0.967670011148272
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.5279186704428867e-05
step: 10, loss: 0.0001488558918936178
step: 20, loss: 4.135813287575729e-05
step: 30, loss: 3.1407307687914e-05
step: 40, loss: 3.0226272428990342e-05
step: 50, loss: 0.0006989146932028234
step: 60, loss: 4.0603084926260635e-05
step: 70, loss: 0.00015455445100087672
step: 80, loss: 5.370835424400866e-05
step: 90, loss: 3.9956721593625844e-05
step: 100, loss: 4.5474916987586766e-05
step: 110, loss: 3.184664456057362e-05
step: 120, loss: 6.985862273722887e-05
step: 130, loss: 0.0004181296972092241
step: 140, loss: 0.00017001802916638553
step: 150, loss: 0.03254682943224907
step: 160, loss: 8.204958430724218e-05
```

```
step: 170, loss: 2.6579447876429185e-05
step: 180, loss: 4.859453474637121e-05
step: 190, loss: 0.00013708621554542333
step: 200, loss: 7.396860746666789e-05
step: 210, loss: 0.010343484580516815
step: 220, loss: 3.931035098503344e-05
step: 230, loss: 0.00011814026947831735
/usr/local/lib/python3.7/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=0.9755011135857461, f1=0.9722530521642618,
best_f1=0.967670011148272
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 [00:07, 338.28it/s]
load_f1 = 0.9797752808988766
real_f1 = 0.9733333333333333
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, 417.00it/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 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
```

372

```
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: 0.7962208390235901
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.45240339636802673
step: 20, loss: 0.5146725177764893
step: 30, loss: 0.4641241133213043
step: 40, loss: 0.4770752489566803
step: 50, loss: 0.4327768385410309
step: 60, loss: 0.49620193243026733
step: 70, loss: 0.18784698843955994
step: 80, loss: 0.12983889877796173
step: 90, loss: 0.17973379790782928
step: 100, loss: 0.266169011592865
step: 110, loss: 0.06025358662009239
step: 120, loss: 0.10096290707588196
step: 130, loss: 0.06180498003959656
step: 140, loss: 0.296170175075531
step: 150, loss: 0.07345665991306305
step: 160, loss: 0.23509329557418823
step: 170, loss: 0.4540939927101135
step: 180, loss: 0.12940283119678497
```

```
step: 190, loss: 0.08643011003732681
step: 200, loss: 0.10837559401988983
step: 210, loss: 0.10077060014009476
step: 220, loss: 0.1304081529378891
step: 230, loss: 0.12837183475494385
step: 240, loss: 0.10688367486000061
step: 250, loss: 0.0881446897983551
step: 260, loss: 0.06964649260044098
step: 270, loss: 0.07021352648735046
step: 280, loss: 0.14160357415676117
step: 290, loss: 0.12100572884082794
step: 300, loss: 0.17226368188858032
step: 310, loss: 0.1686905473470688
step: 320, loss: 0.1602027416229248
step: 330, loss: 0.18323291838169098
step: 340, loss: 0.2626543939113617
step: 350, loss: 0.166444331407547
step: 360, loss: 0.09450706839561462
step: 370, loss: 0.14333681762218475
step: 380, loss: 0.09005676954984665
step: 390, loss: 0.1769273430109024
step: 400, loss: 0.030528325587511063
step: 410, loss: 0.11689458787441254
step: 420, loss: 0.03779524937272072
step: 430, loss: 0.08277522027492523
step: 440, loss: 0.2795688509941101
step: 450, loss: 0.04352149739861488
step: 460, loss: 0.03861450403928757
step: 470, loss: 0.2723231613636017
step: 480, loss: 0.23819302022457123
step: 490, loss: 0.18933254480361938
step: 500, loss: 0.03552057966589928
step: 510, loss: 0.07346659898757935
step: 520, loss: 0.12829819321632385
step: 530, loss: 0.2761383056640625
/usr/local/lib/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=0.8999538958045182, f1=0.8901200369344413,
best_f1=0.8901200369344413
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.14251986145973206
```

```
step: 10, loss: 0.062065042555332184
step: 20, loss: 0.13641692698001862
step: 30, loss: 0.09132867306470871
step: 40, loss: 0.029878711327910423
step: 50, loss: 0.08046898245811462
step: 60, loss: 0.13707831501960754
step: 70, loss: 0.19255860149860382
step: 80, loss: 0.011465254239737988
step: 90, loss: 0.04843004420399666
step: 100, loss: 0.3062833249568939
step: 110, loss: 0.08812291920185089
step: 120, loss: 0.07916849851608276
step: 130, loss: 0.05000113323330879
step: 140, loss: 0.02877311035990715
step: 150, loss: 0.13320337235927582
step: 160, loss: 0.23096147179603577
step: 170, loss: 0.024058327078819275
step: 180, loss: 0.029267892241477966
step: 190, loss: 0.08212241530418396
step: 200, loss: 0.1698990762233734
step: 210, loss: 0.05398376286029816
step: 220, loss: 0.16005678474903107
step: 230, loss: 0.16152329742908478
step: 240, loss: 0.11311998218297958
step: 250, loss: 0.06328899413347244
step: 260, loss: 0.03432203829288483
step: 270, loss: 0.14704681932926178
step: 280, loss: 0.18631580471992493
step: 290, loss: 0.1001850962638855
step: 300, loss: 0.07798206806182861
step: 310, loss: 0.13322687149047852
step: 320, loss: 0.13543754816055298
step: 330, loss: 0.06500188261270523
step: 340, loss: 0.015757372602820396
step: 350, loss: 0.08654693514108658
step: 360, loss: 0.15829506516456604
step: 370, loss: 0.027066944167017937
step: 380, loss: 0.10826614499092102
step: 390, loss: 0.11051066219806671
step: 400, loss: 0.08767487108707428
step: 410, loss: 0.008579627610743046
step: 420, loss: 0.05316021665930748
step: 430, loss: 0.057026248425245285
step: 440, loss: 0.017977671697735786
step: 450, loss: 0.130628764629364
step: 460, loss: 0.16051341593265533
step: 470, loss: 0.043825797736644745
step: 480, loss: 0.3169213533401489
```

```
step: 490, loss: 0.1049647256731987
step: 500, loss: 0.0449383519589901
step: 510, loss: 0.1493045687675476
step: 520, loss: 0.06762092560529709
step: 530, loss: 0.17019535601139069
/usr/local/lib/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=0.9057318806252961, f1=0.890057361376673,
best_f1=0.890057361376673
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04110715165734291
step: 10, loss: 0.19656851887702942
step: 20, loss: 0.11372288316488266
step: 30, loss: 0.2728493809700012
step: 40, loss: 0.03351106867194176
step: 50, loss: 0.06234405189752579
step: 60, loss: 0.012086108326911926
step: 70, loss: 0.006936792284250259
step: 80, loss: 0.22105184197425842
step: 90, loss: 0.20003941655158997
step: 100, loss: 0.0209360271692276
step: 110, loss: 0.06888946145772934
step: 120, loss: 0.09694348275661469
step: 130, loss: 0.06270432472229004
step: 140, loss: 0.02233877032995224
step: 150, loss: 0.0785071924328804
step: 160, loss: 0.0248276237398386
step: 170, loss: 0.03848734498023987
step: 180, loss: 0.028406841680407524
step: 190, loss: 0.00473262881860137
step: 200, loss: 0.05245091766119003
step: 210, loss: 0.06774582713842392
step: 220, loss: 0.015371538698673248
step: 230, loss: 0.010242470540106297
step: 240, loss: 0.008870039135217667
step: 250, loss: 0.030476344749331474
step: 260, loss: 0.023625267669558525
step: 270, loss: 0.003828353714197874
step: 280, loss: 0.07936891913414001
step: 290, loss: 0.1755920648574829
step: 300, loss: 0.0477004311978817
```

```
step: 310, loss: 0.23543108999729156
step: 320, loss: 0.06601742655038834
step: 330, loss: 0.015656061470508575
step: 340, loss: 0.00805297028273344
step: 350, loss: 0.1305793970823288
step: 360, loss: 0.008107698522508144
step: 370, loss: 0.015508070588111877
step: 380, loss: 0.06352129578590393
step: 390, loss: 0.04187529906630516
step: 400, loss: 0.02021603286266327
step: 410, loss: 0.09517636895179749
step: 420, loss: 0.06576332449913025
step: 430, loss: 0.036433879286050797
step: 440, loss: 0.11432290822267532
step: 450, loss: 0.1734548658132553
step: 460, loss: 0.143052026629944794
step: 470, loss: 0.01315336674451828
step: 480, loss: 0.015126921236515045
step: 490, loss: 0.02473350614309311
step: 500, loss: 0.06693438440561295
step: 510, loss: 0.009074845351278782
step: 520, loss: 0.010306923650205135
step: 530, loss: 0.010773374699056149
/usr/local/lib/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=0.9063670411985018, f1=0.9027454630060493,
best_f1=0.9027454630060493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0073103224858641624
step: 10, loss: 0.003320542396977544
step: 20, loss: 0.1362529993057251
step: 30, loss: 0.06551088392734528
step: 40, loss: 0.056882183998823166
step: 50, loss: 0.02853415720164776
step: 60, loss: 0.004501865711063147
step: 70, loss: 0.005438623484224081
step: 80, loss: 0.013039987534284592
step: 90, loss: 0.10791052877902985
step: 100, loss: 0.052274417132139206
step: 110, loss: 0.007742627523839474
step: 120, loss: 0.03046443685889244
```

```
step: 130, loss: 0.015626147389411926
step: 140, loss: 0.017594151198863983
step: 150, loss: 0.053912024945020676
step: 160, loss: 0.006508626043796539
step: 170, loss: 0.003856698516756296
step: 180, loss: 0.03766312450170517
step: 190, loss: 0.02809026651084423
step: 200, loss: 0.012429637834429741
step: 210, loss: 0.06681369990110397
step: 220, loss: 0.024209247902035713
step: 230, loss: 0.32829877734184265
step: 240, loss: 0.015343692153692245
step: 250, loss: 0.0037957525346428156
step: 260, loss: 0.08274944871664047
step: 270, loss: 0.18230803310871124
step: 280, loss: 0.0021089077927172184
step: 290, loss: 0.060283221304416656
step: 300, loss: 0.003284615697339177
step: 310, loss: 0.017286261543631554
step: 320, loss: 0.07137191295623779
step: 330, loss: 0.18019449710845947
step: 340, loss: 0.15743137896060944
step: 350, loss: 0.014208676293492317
step: 360, loss: 0.06381186097860336
step: 370, loss: 0.036725401878356934
step: 380, loss: 0.029105307534337044
step: 390, loss: 0.05147971957921982
step: 400, loss: 0.07453811168670654
step: 410, loss: 0.08119925111532211
step: 420, loss: 0.04888863116502762
step: 430, loss: 0.014295362867414951
step: 440, loss: 0.0682268813252449
step: 450, loss: 0.22595623135566711
step: 460, loss: 0.003704611910507083
step: 470, loss: 0.02559526264667511
step: 480, loss: 0.020226871594786644
step: 490, loss: 0.034705307334661484
step: 500, loss: 0.01814093181490898
step: 510, loss: 0.14736175537109375
step: 520, loss: 0.11593315005302429
step: 530, loss: 0.0038595865480601788
/usr/local/lib/python3.7/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=0.9023474178403756, f1=0.8940737284181055,
best_f1=0.9027454630060493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02698279730975628
step: 10, loss: 0.02216852642595768
step: 20, loss: 0.023715874180197716
step: 30, loss: 0.00909302569925785
step: 40, loss: 0.11720822751522064
step: 50, loss: 0.09556514769792557
step: 60, loss: 0.0036565472837537527
step: 70, loss: 0.06978994607925415
step: 80, loss: 0.006877817213535309
step: 90, loss: 0.002764668082818389
step: 100, loss: 0.0015421593561768532
step: 110, loss: 0.0028273207135498524
step: 120, loss: 0.0018124980852007866
step: 130, loss: 0.003443677444010973
step: 140, loss: 0.006465388461947441
step: 150, loss: 0.15316110849380493
step: 160, loss: 0.00596837792545557
step: 170, loss: 0.009327967651188374
step: 180, loss: 0.0019358759745955467
step: 190, loss: 0.007297426927834749
step: 200, loss: 0.013115284033119678
step: 210, loss: 0.051278356462717056
step: 220, loss: 0.0018647693796083331
step: 230, loss: 0.0012189610861241817
step: 240, loss: 0.04734230414032936
step: 250, loss: 0.0008721953490749002
step: 260, loss: 0.03878113254904747
step: 270, loss: 0.03758919611573219
step: 280, loss: 0.0251216851748314
step: 290, loss: 0.001664129551500082
step: 300, loss: 0.16630595922470093
step: 310, loss: 0.006214917171746492
step: 320, loss: 0.00994543544948101
step: 330, loss: 0.021349657326936722
step: 340, loss: 0.017026619985699654
step: 350, loss: 0.020230643451213837
step: 360, loss: 0.06483243405818939
step: 370, loss: 0.008596659637987614
step: 380, loss: 0.06367364525794983
step: 390, loss: 0.01079567614942789
step: 400, loss: 0.012964032590389252
step: 410, loss: 0.003305052639916539
step: 420, loss: 0.005487170536071062
step: 430, loss: 0.0015695802867412567
step: 440, loss: 0.03652261570096016
step: 450, loss: 0.067457415163517
step: 460, loss: 0.020720968022942543
```

```
step: 470, loss: 0.008232245221734047
step: 480, loss: 0.0012066466733813286
step: 490, loss: 0.0016438113525509834
step: 500, loss: 0.01408644299954176
step: 510, loss: 0.06012646108865738
step: 520, loss: 0.006216266192495823
step: 530, loss: 0.005175726022571325
/usr/local/lib/python3.7/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=0.9031665901789813, f1=0.9010584445467096,
best_f1=0.9027454630060493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004680450540035963
step: 10, loss: 0.005016134586185217
step: 20, loss: 0.00032884228858165443
step: 30, loss: 0.000684837403241545
step: 40, loss: 0.005414573475718498
step: 50, loss: 0.012345099821686745
step: 60, loss: 0.0005644093616865575
step: 70, loss: 0.056608397513628006
step: 80, loss: 0.03503487631678581
step: 90, loss: 0.019852666184306145
step: 100, loss: 0.01911591738462448
step: 110, loss: 0.013340656645596027
step: 120, loss: 0.013264870271086693
step: 130, loss: 0.020321620628237724
step: 140, loss: 0.00477188965305686
step: 150, loss: 0.005033575929701328
step: 160, loss: 0.04268990829586983
step: 170, loss: 0.00069779891055077731
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 180, loss: 0.004151505883783102
step: 190, loss: 0.0005564783350564539
step: 200, loss: 0.001245860941708088
step: 210, loss: 0.012972329743206501
step: 220, loss: 0.0006872020312584937
step: 230, loss: 0.006425303407013416
step: 240, loss: 0.12209590524435043
step: 250, loss: 0.013733088970184326
step: 260, loss: 0.0030146820936352015
step: 270, loss: 0.0029966377187520266
step: 280, loss: 0.02782120183110237
step: 290, loss: 0.006555716507136822
step: 300, loss: 0.01898670196533203
step: 310, loss: 0.03805257007479668
```

```
step: 320, loss: 0.20741751790046692
step: 330, loss: 0.00872295256704092
step: 340, loss: 0.013545106165111065
step: 350, loss: 0.12325219064950943
step: 360, loss: 0.003114728257060051
step: 370, loss: 0.011891528032720089
step: 380, loss: 0.004123717080801725
step: 390, loss: 0.1396486461162567
step: 400, loss: 0.0011583009036257863
step: 410, loss: 0.0012581013143062592
step: 420, loss: 0.10716002434492111
step: 430, loss: 0.04217642545700073
step: 440, loss: 0.08671144396066666
step: 450, loss: 0.014297859743237495
step: 460, loss: 0.04002038389444351
step: 470, loss: 0.12639039754867554
step: 480, loss: 0.005964160431176424
step: 490, loss: 0.014423368498682976
step: 500, loss: 0.008271818980574608
step: 510, loss: 0.0007389733218587935
step: 520, loss: 0.03153461590409279
step: 530, loss: 0.0017379054334014654
/usr/local/lib/python3.7/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=0.8995305164319248, f1=0.8900424728645587,
best_f1=0.9027454630060493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06159514933824539
step: 10, loss: 0.016702065244317055
step: 20, loss: 0.002920126775279641
step: 30, loss: 0.04378432780504227
step: 40, loss: 0.005719680339097977
step: 50, loss: 0.008100222796201706
step: 60, loss: 0.011897657066583633
step: 70, loss: 0.02185901068150997
step: 80, loss: 0.0157729871571064
step: 90, loss: 0.004510801285505295
step: 100, loss: 0.005952063482254744
step: 110, loss: 0.0033148303627967834
step: 120, loss: 0.021519910544157028
step: 130, loss: 0.07720110565423965
step: 140, loss: 0.09719906002283096
step: 150, loss: 0.0033562686294317245
step: 160, loss: 0.0002952925569843501
step: 170, loss: 0.0008321120985783637
```

```
step: 180, loss: 0.006510838866233826
step: 190, loss: 0.0022852581460028887
step: 200, loss: 0.002340524224564433
step: 210, loss: 0.0036753315944224596
step: 220, loss: 0.0012753314804285765
step: 230, loss: 0.0019734669476747513
step: 240, loss: 0.008456462062895298
step: 250, loss: 0.005370312836021185
step: 260, loss: 0.001996066654101014
step: 270, loss: 0.0023446609266102314
step: 280, loss: 0.006191902328282595
step: 290, loss: 0.0021884306333959103
step: 300, loss: 0.002486068056896329
step: 310, loss: 0.0015791055047884583
step: 320, loss: 0.014707997441291809
step: 330, loss: 0.13361482322216034
step: 340, loss: 0.023520637303590775
step: 350, loss: 0.00040344733861275017
step: 360, loss: 0.003710536053404212
step: 370, loss: 0.08191792666912079
step: 380, loss: 0.01976863667368889
step: 390, loss: 0.0024652364663779736
step: 400, loss: 0.05233769491314888
step: 410, loss: 0.007110987789928913
step: 420, loss: 0.0006584431394003332
step: 430, loss: 0.0006233700551092625
step: 440, loss: 0.0035204661544412374
step: 450, loss: 0.0008708798559382558
step: 460, loss: 0.03820662945508957
step: 470, loss: 0.25860655307769775
step: 480, loss: 0.0009085287456400692
step: 490, loss: 0.00238681398332119
step: 500, loss: 0.0003937946748919785
step: 510, loss: 0.005973976105451584
step: 520, loss: 0.18858025968074799
step: 530, loss: 0.0037424226757138968
/usr/local/lib/python3.7/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=0.896030245746692, f1=0.8959700093720713,
best_f1=0.9027454630060493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003045530989766121
step: 10, loss: 0.08257834613323212
step: 20, loss: 0.004759186413139105
step: 30, loss: 0.001530241221189499
```

```
step: 40, loss: 0.007739974185824394
step: 50, loss: 0.000577569764573127
step: 60, loss: 0.0003422510053496808
step: 70, loss: 0.0013914515729993582
step: 80, loss: 0.0003819323319476098
step: 90, loss: 0.00015888364578131586
step: 100, loss: 0.07398992031812668
step: 110, loss: 0.0009453541715629399
step: 120, loss: 0.011630724184215069
step: 130, loss: 0.001715713762678206
step: 140, loss: 0.00010081783693749458
step: 150, loss: 0.0004342882602941245
step: 160, loss: 0.0005539903650060296
step: 170, loss: 0.00020794356532860547
step: 180, loss: 0.0002515890228096396
step: 190, loss: 0.0012573577696457505
step: 200, loss: 0.0038430076092481613
step: 210, loss: 0.0012138454476371408
step: 220, loss: 0.0035013232845813036
step: 230, loss: 0.00019136314222123474
step: 240, loss: 0.11232120543718338
step: 250, loss: 0.0010306148324161768
step: 260, loss: 0.13777075707912445
step: 270, loss: 0.000493975356221199
step: 280, loss: 0.01423371396958828
step: 290, loss: 0.0020494554191827774
step: 300, loss: 0.003177644219249487
step: 310, loss: 0.011287147179245949
step: 320, loss: 0.0011347121326252818
step: 330, loss: 0.047149986028671265
step: 340, loss: 0.00024079506692942232
step: 350, loss: 0.02425096370279789
step: 360, loss: 0.01814117096364498
step: 370, loss: 0.0049249534495174885
step: 380, loss: 0.002232924336567521
step: 390, loss: 0.0009366340236738324
step: 400, loss: 0.007889585569500923
step: 410, loss: 0.01421884074807167
step: 420, loss: 0.00029136062948964536
step: 430, loss: 0.00034823009627871215
step: 440, loss: 0.0022088238038122654
step: 450, loss: 0.005313721019774675
step: 460, loss: 0.0032594462390989065
step: 470, loss: 0.0009673749445937574
step: 480, loss: 0.0009797756792977452
step: 490, loss: 0.00027085954206995666
step: 500, loss: 0.0017327879322692752
step: 510, loss: 0.0034588815178722143
```

```
step: 520, loss: 0.00036204938078299165
step: 530, loss: 0.0003751624608412385
/usr/local/lib/python3.7/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=0.8993039443155452, f1=0.8966480446927374,
best_f1=0.9027454630060493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009623699006662452
step: 10, loss: 0.003909677267074585
step: 20, loss: 0.001201748731546104
step: 30, loss: 0.000787049881182611
step: 40, loss: 0.0024388409219682217
step: 50, loss: 0.0008647070499137044
step: 60, loss: 0.0009985575452446938
step: 70, loss: 0.008809780701994896
step: 80, loss: 0.00559203838929534
step: 90, loss: 0.0022243044804781675
step: 100, loss: 0.0014331587590277195
step: 110, loss: 0.004981242585927248
step: 120, loss: 0.1039477288722992
step: 130, loss: 0.0019815790001302958
step: 140, loss: 0.00084579276153811777
step: 150, loss: 0.000847615476232022
step: 160, loss: 0.00018157184240408242
step: 170, loss: 0.00016663111455272883
step: 180, loss: 0.0003769753675442189
step: 190, loss: 0.015415113419294357
step: 200, loss: 0.0017054729396477342
step: 210, loss: 0.00022644751879852265
step: 220, loss: 0.09579765051603317
step: 230, loss: 0.004461487289518118
step: 240, loss: 0.04223603382706642
step: 250, loss: 9.439195855520666e-05
step: 260, loss: 0.00032717431895434856
step: 270, loss: 0.00361329922452569
step: 280, loss: 0.013304183259606361
step: 290, loss: 0.0011684661731123924
step: 300, loss: 0.0013977650087326765
step: 310, loss: 0.00013458965986501426
step: 320, loss: 0.00023328552197199315
step: 330, loss: 0.0027764562983065844
step: 340, loss: 0.00021685179672203958
step: 350, loss: 9.161206253338605e-05
step: 360, loss: 0.0038089274894446135
step: 370, loss: 0.0006516706780530512
```

```
step: 380, loss: 0.002999226562678814
step: 390, loss: 0.0044676694087684155
step: 400, loss: 0.004484476521611214
step: 410, loss: 0.004421387333422899
step: 420, loss: 0.0020990725606679916
step: 430, loss: 7.70676342654042e-05
step: 440, loss: 0.0011764310766011477
step: 450, loss: 0.01959751546382904
step: 460, loss: 0.0004220456175971776
step: 470, loss: 0.00029799120966345507
step: 480, loss: 0.04287257418036461
step: 490, loss: 0.00024268559354823083
step: 500, loss: 0.000351451279129833
step: 510, loss: 9.858674457063898e-05
step: 520, loss: 0.00011446340067777783
step: 530, loss: 0.06296223402023315
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 8192.0
/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 9: dev_f1=0.8993536472760849, f1=0.900875979714154,
best_f1=0.9027454630060493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.22814083099365234
step: 10, loss: 9.934745321515948e-05
step: 20, loss: 0.000296841433737427
step: 30, loss: 0.00024823661078698933
step: 40, loss: 0.0016113148303702474
step: 50, loss: 0.0002508573525119573
step: 60, loss: 0.000552161131054163
step: 70, loss: 0.0008827560814097524
step: 80, loss: 0.2449568808078766
step: 90, loss: 0.00023125388543121517
step: 100, loss: 0.0003345434961374849
step: 110, loss: 0.0015703316312283278
step: 120, loss: 0.0002788534911815077
step: 130, loss: 9.925048652803525e-05
step: 140, loss: 0.027002466842532158
step: 150, loss: 9.300248348154128e-05
step: 160, loss: 0.00032843052758835256
step: 170, loss: 0.0003606760292313993
step: 180, loss: 0.010442148894071579
step: 190, loss: 0.005717063322663307
step: 200, loss: 0.006540337111800909
step: 210, loss: 0.0009406742756254971
step: 220, loss: 0.06795637309551239
```

```
step: 230, loss: 0.0001831122353905812
step: 240, loss: 0.0007049958221614361
step: 250, loss: 0.00013755467080045491
step: 260, loss: 0.000264503265498206
step: 270, loss: 0.00010082965309266001
step: 280, loss: 8.345681999344379e-05
step: 290, loss: 6.887525523779914e-05
step: 300, loss: 9.9273442174308e-05
step: 310, loss: 0.0031929723918437958
step: 320, loss: 0.0002680261386558435
step: 330, loss: 0.0006969129317440093
step: 340, loss: 0.021139970049262047
step: 350, loss: 0.00015362666454166174
step: 360, loss: 0.00014584125892724842
step: 370, loss: 0.0009510915260761976
step: 380, loss: 0.004967842251062393
step: 390, loss: 0.00023705873172730207
step: 400, loss: 0.00016695550584699959
step: 410, loss: 0.00022706396703142673
step: 420, loss: 0.010191482491791248
step: 430, loss: 0.0003153861325699836
step: 440, loss: 5.364164462662302e-05
step: 450, loss: 0.0033098275307565928
step: 460, loss: 0.0008879936649464071
step: 470, loss: 0.0004610349715221673
step: 480, loss: 0.00023072260955814272
step: 490, loss: 0.030857935547828674
step: 500, loss: 0.001063888892531395
step: 510, loss: 0.0007269468624144793
step: 520, loss: 0.00189517333637923
step: 530, loss: 0.002592496806755662
/usr/local/lib/python3.7/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=0.9022346368715083, f1=0.8997214484679666,
best_f1=0.9027454630060493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002816690248437226
step: 10, loss: 0.000609991024248302
step: 20, loss: 0.00013929937267675996
step: 30, loss: 0.0632651224732399
step: 40, loss: 0.01081408467143774
step: 50, loss: 0.0003473739488981664
step: 60, loss: 0.0038319791201502085
step: 70, loss: 0.00010288324119755998
step: 80, loss: 0.01572365127503872
```

```
step: 90, loss: 0.0008861118112690747
step: 100, loss: 0.0007493494194932282
step: 110, loss: 0.017988869920372963
step: 120, loss: 0.04425783455371857
step: 130, loss: 0.0009512359974905849
step: 140, loss: 0.0005351672298274934
step: 150, loss: 6.809618935221806e-05
step: 160, loss: 0.0014821148943156004
step: 170, loss: 0.0033869340550154448
step: 180, loss: 0.0004974076291546226
step: 190, loss: 0.004467319697141647
step: 200, loss: 0.001739105093292892
step: 210, loss: 0.0008785440004430711
step: 220, loss: 0.00033569676452316344
step: 230, loss: 0.0004337869177106768
step: 240, loss: 0.00029975306824781
step: 250, loss: 0.001016576192341745
step: 260, loss: 0.0003205068060196936
step: 270, loss: 0.0028547008987516165
step: 280, loss: 0.00014249482774175704
step: 290, loss: 0.0003490860399324447
step: 300, loss: 0.06675835698843002
step: 310, loss: 0.032106198370456696
step: 320, loss: 0.005010400898754597
step: 330, loss: 0.02156633511185646
step: 340, loss: 0.0002661525213625282
step: 350, loss: 0.00640815868973732
step: 360, loss: 0.000878495629876852
step: 370, loss: 0.00026784130022861063
step: 380, loss: 0.00024920038413256407
step: 390, loss: 0.0035464761313050985
step: 400, loss: 7.960290531627834e-05
step: 410, loss: 0.0003668582357931882
step: 420, loss: 0.0017498470842838287
step: 430, loss: 0.0001318879658356309
step: 440, loss: 0.001962395152077079
step: 450, loss: 0.00010190177999902517
step: 460, loss: 0.0025064500514417887
step: 470, loss: 0.0022439288441091776
step: 480, loss: 0.012874843552708626
step: 490, loss: 0.018947875127196312
step: 500, loss: 7.619248935952783e-05
step: 510, loss: 0.0008198341820389032
step: 520, loss: 0.00018487274064682424
step: 530, loss: 0.0031106751412153244
/usr/local/lib/python3.7/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=0.9031665901789813, f1=0.8956999085086915,
best_f1=0.9027454630060493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008558580884709954
step: 10, loss: 8.335168240591884e-05
step: 20, loss: 2.8319129341980442e-05
step: 30, loss: 0.004737495444715023
step: 40, loss: 0.0001271959044970572
step: 50, loss: 0.11833216995000839
step: 60, loss: 0.00037310938932932913
step: 70, loss: 0.05426545813679695
step: 80, loss: 0.00019941217033192515
step: 90, loss: 0.0007597083458676934
step: 100, loss: 0.00042929392657242715
step: 110, loss: 0.004128274042159319
step: 120, loss: 0.10215316712856293
step: 130, loss: 0.0029541479889303446
step: 140, loss: 9.960451279766858e-05
step: 150, loss: 0.0025631021708250046
step: 160, loss: 6.997343734838068e-05
step: 170, loss: 0.0030730292201042175
step: 180, loss: 6.0858583310619e-05
step: 190, loss: 9.377855167258531e-05
step: 200, loss: 0.0017337003955617547
step: 210, loss: 0.005183802917599678
step: 220, loss: 0.013745909556746483
step: 230, loss: 0.009592709131538868
step: 240, loss: 0.008591708727180958
step: 250, loss: 0.004445330239832401
step: 260, loss: 0.0008742101490497589
step: 270, loss: 0.0005232832627370954
step: 280, loss: 0.004159536212682724
step: 290, loss: 0.0007537738420069218
step: 300, loss: 0.0744992196559906
step: 310, loss: 0.002382401144132018
step: 320, loss: 0.041576024144887924
step: 330, loss: 0.013269046321511269
step: 340, loss: 0.002385195577517152
step: 350, loss: 0.010348541662096977
step: 360, loss: 0.0001840592740336433
step: 370, loss: 0.0001110283046728
step: 380, loss: 0.00016440583567600697
step: 390, loss: 5.548700210056268e-05
step: 400, loss: 0.0003175109450239688
step: 410, loss: 0.00017780449707061052
step: 420, loss: 0.005644018761813641
```

```
step: 430, loss: 0.00035493462928570807
step: 440, loss: 0.07341079413890839
step: 450, loss: 0.0003999117761850357
step: 460, loss: 0.003442251356318593
step: 470, loss: 0.0016327629564329982
step: 480, loss: 0.0023110040929168463
step: 490, loss: 5.079768016003072e-05
step: 500, loss: 0.0008290472324006259
step: 510, loss: 9.764944115886465e-05
step: 520, loss: 0.015944113954901695
step: 530, loss: 0.0003509200469125062
/usr/local/lib/python3.7/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=0.8959107806691451, f1=0.8921795465062471,
best_f1=0.9027454630060493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0019134802278131247
step: 10, loss: 0.00010535571345826611
step: 20, loss: 0.013235378079116344
step: 30, loss: 0.012795322574675083
step: 40, loss: 8.926322334446013e-05
step: 50, loss: 0.0034928354434669018
step: 60, loss: 0.00015719622024334967
step: 70, loss: 5.967697507003322e-05
step: 80, loss: 0.0005837525241076946
step: 90, loss: 0.004209986422210932
step: 100, loss: 7.034843292785808e-05
step: 110, loss: 0.0002091568021569401
step: 120, loss: 0.0005085683660581708
step: 130, loss: 0.00015877251280471683
step: 140, loss: 0.0001911168801598251
step: 150, loss: 0.0009592222049832344
step: 160, loss: 0.004294654820114374
step: 170, loss: 0.0007899185875430703
step: 180, loss: 0.0011278268648311496
step: 190, loss: 0.0011807173723354936
step: 200, loss: 0.000759391812607646
step: 210, loss: 0.005586729384958744
step: 220, loss: 0.00015211119898594916
step: 230, loss: 8.734581933822483e-05
step: 240, loss: 0.00013319849676918238
step: 250, loss: 0.02174890600144863
step: 260, loss: 0.0004517019260674715
step: 270, loss: 0.00041162516572512686
step: 280, loss: 0.0007169835735112429
```

```
step: 290, loss: 0.0025090330746024847
step: 300, loss: 0.002898824168369174
step: 310, loss: 0.008598837070167065
step: 320, loss: 0.00016698517720215023
step: 330, loss: 0.00024867584579624236
step: 340, loss: 0.0003830324567388743
step: 350, loss: 0.0003825822495855391
step: 360, loss: 0.006344336550682783
step: 370, loss: 0.0006658190977759659
step: 380, loss: 0.00011705694487318397
step: 390, loss: 0.00057525176089257
step: 400, loss: 0.0001865230151452124
step: 410, loss: 0.00029313290724530816
step: 420, loss: 0.00016064110968727618
step: 430, loss: 0.0007153185433708131
step: 440, loss: 0.003133901860564947
step: 450, loss: 0.3841778635978699
step: 460, loss: 0.00012029457138851285
step: 470, loss: 0.0034258048981428146
step: 480, loss: 0.00038582313572987914
step: 490, loss: 0.0004142780671827495
step: 500, loss: 0.00039281678618863225
step: 510, loss: 7.98778492026031e-05
step: 520, loss: 0.00029190070927143097
step: 530, loss: 0.0003061641473323107
/usr/local/lib/python3.7/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=0.89749430523918, f1=0.8964888280893752,
best_f1=0.9027454630060493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010381718166172504
step: 10, loss: 0.005382867064327002
step: 20, loss: 0.0006358257378451526
step: 30, loss: 0.0007940993527881801
step: 40, loss: 0.0016451855190098286
step: 50, loss: 0.0046710846945643425
step: 60, loss: 5.6487639085389674e-05
step: 70, loss: 0.0013404550263658166
step: 80, loss: 0.00011639976582955569
step: 90, loss: 0.0001696411200100556
step: 100, loss: 0.0028735478408634663
step: 110, loss: 0.00016930473793763667
step: 120, loss: 0.00014836870832368731
step: 130, loss: 0.00029695749981328845
step: 140, loss: 0.02551920898258686
```

```
step: 150, loss: 0.00032053780159913003
step: 160, loss: 0.00018889919738285244
step: 170, loss: 0.001594259636476636
step: 180, loss: 0.00010874262807192281
step: 190, loss: 9.957711154129356e-05
step: 200, loss: 0.0007791888201609254
step: 210, loss: 0.0002469445753376931
step: 220, loss: 8.848016295814887e-05
step: 230, loss: 0.001371111604385078
step: 240, loss: 0.00010087182454299182
step: 250, loss: 0.0007037131581455469
step: 260, loss: 4.499174247030169e-05
step: 270, loss: 0.0037434531841427088
step: 280, loss: 0.00028194356127642095
step: 290, loss: 0.0013555530458688736
step: 300, loss: 0.000976675539277494
step: 310, loss: 0.0006641946965828538
step: 320, loss: 4.353599069872871e-05
step: 330, loss: 0.01984735205769539
step: 340, loss: 0.00018766173161566257
step: 350, loss: 0.0025836925487965345
step: 360, loss: 0.004883033689111471
step: 370, loss: 0.0028828687500208616
step: 380, loss: 0.027705825865268707
step: 390, loss: 0.013630061410367489
step: 400, loss: 0.000116694689495489
step: 410, loss: 0.00029955743229947984
step: 420, loss: 0.0017881713574752212
step: 430, loss: 0.0005890716565772891
step: 440, loss: 0.0013277486432343721
step: 450, loss: 0.00010998254583682865
step: 460, loss: 0.035446155816316605
step: 470, loss: 0.0006042335298843682
step: 480, loss: 0.0001956464839167893
step: 490, loss: 0.0006463929894380271
step: 500, loss: 0.0023440271615982056
step: 510, loss: 0.0002024737186729908
step: 520, loss: 0.0001031463107210584
step: 530, loss: 9.081820462597534e-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=0.8952205882352942, f1=0.897119341563786,
best_f1=0.9027454630060493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00015560330939479172
```

```
step: 10, loss: 7.000496407272294e-05
step: 20, loss: 0.0006924147019162774
step: 30, loss: 0.0001886104728328064
step: 40, loss: 0.0005382175440900028
step: 50, loss: 0.0002611281524877995
step: 60, loss: 0.00011797479237429798
step: 70, loss: 0.0004875595332123339
step: 80, loss: 0.00319140893407166
step: 90, loss: 0.0002461292024261724
step: 100, loss: 7.185762660810724e-05
step: 110, loss: 0.3063594400882721
step: 120, loss: 0.001268164487555623
step: 130, loss: 0.0003203761298209429
step: 140, loss: 4.0376799006480724e-05
step: 150, loss: 0.004333821125328541
step: 160, loss: 0.0006107601220719516
step: 170, loss: 0.00033380003878846765
step: 180, loss: 0.0001014614463201724
step: 190, loss: 0.022832509130239487
step: 200, loss: 0.00020683756156358868
step: 210, loss: 0.0009635832393541932
step: 220, loss: 0.0006635875906795263
step: 230, loss: 0.00014557030226569623
step: 240, loss: 0.00012939066800754517
step: 250, loss: 0.011845709756016731
step: 260, loss: 0.00011370598804205656
step: 270, loss: 0.00023125008738134056
step: 280, loss: 0.00023597344988957047
step: 290, loss: 0.014116568490862846
step: 300, loss: 0.0066685727797448635
step: 310, loss: 0.0014746938832104206
step: 320, loss: 0.00031953310826793313
step: 330, loss: 0.009405773133039474
step: 340, loss: 0.0054845078848302364
step: 350, loss: 0.0004086285480298102
step: 360, loss: 0.0002912074269261211
step: 370, loss: 6.023043533787131e-05
step: 380, loss: 3.439445936237462e-05
step: 390, loss: 0.00020465074339881540
step: 400, loss: 8.059047831920907e-05
step: 410, loss: 0.005884159822016954
step: 420, loss: 6.80003286106512e-05
step: 430, loss: 3.0196413717931136e-05
step: 440, loss: 0.0065803928300738335
step: 450, loss: 0.012271305546164513
step: 460, loss: 7.388190715573728e-05
step: 470, loss: 0.00015665244427509606
step: 480, loss: 0.006226339843124151
```

```
step: 490, loss: 0.035045791417360306
step: 500, loss: 7.806938810972497e-05
step: 510, loss: 0.00014348413969855756
step: 520, loss: 0.007989509031176567
step: 530, loss: 0.0014587517362087965
/usr/local/lib/python3.7/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=0.8979963570127504, f1=0.8953594176524113,
best_f1=0.9027454630060493
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.")
5742it [00:16, 349.26it/s]
load_f1 = 0.9010270774976656
real_f1 = 0.8997668997668998
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, 388.12it/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 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: 0.8441272377967834
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.06626336627653122
step: 20, loss: 0.38605162501335144
step: 30, loss: 0.3739221394062042
step: 40, loss: 0.5057337284088135
step: 50, loss: 0.31204909086227417
step: 60, loss: 0.36970624327659607
step: 70, loss: 0.27907049655914307
step: 80, loss: 0.29751619696617126
step: 90, loss: 0.4204917550086975
step: 100, loss: 0.19495804607868195
step: 110, loss: 0.3003688454627991
step: 120, loss: 0.25707462430000305
step: 130, loss: 0.24583424627780914
step: 140, loss: 0.3085329830646515
step: 150, loss: 0.32011765241622925
step: 160, loss: 0.25933703780174255
step: 170, loss: 0.22035370767116547
step: 180, loss: 0.18370015919208527
step: 190, loss: 0.26661545038223267
step: 200, loss: 0.18901363015174866
step: 210, loss: 0.4963773190975189
```

```
/usr/local/lib/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=0.3703703703703704, f1=0.36538461538461536,
best_f1=0.36538461538461536
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10539943724870682
step: 10, loss: 0.07897390425205231
step: 20, loss: 0.290034681558609
step: 30, loss: 0.16430366039276123
step: 40, loss: 0.05375111103057861
step: 50, loss: 0.22843486070632935
step: 60, loss: 0.14982455968856812
step: 70, loss: 0.2739974558353424
step: 80, loss: 0.28957879543304443
step: 90, loss: 0.13701632618904114
step: 100, loss: 0.17746712267398834
step: 110, loss: 0.06645005941390991
step: 120, loss: 0.20741496980190277
step: 130, loss: 0.2555668354034424
step: 140, loss: 0.28611981868743896
step: 150, loss: 0.3140665292739868
step: 160, loss: 0.1809387505054474
step: 170, loss: 0.18532200157642365
step: 180, loss: 0.3818863332271576
step: 190, loss: 0.21687495708465576
step: 200, loss: 0.2318144589662552
step: 210, loss: 0.3657529950141907
/usr/local/lib/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=0.4532374100719424, f1=0.5105633802816901,
best_f1=0.5105633802816901
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2741549015045166
step: 10, loss: 0.2737467885017395
```

```
step: 20, loss: 0.4599064588546753
step: 30, loss: 0.07365485280752182
step: 40, loss: 0.18326368927955627
step: 50, loss: 0.19705244898796082
step: 60, loss: 0.2920210361480713
step: 70, loss: 0.30249860882759094
step: 80, loss: 0.0876728817820549
step: 90, loss: 0.13426613807678223
step: 100, loss: 0.13766898214817047
step: 110, loss: 0.2725963592529297
step: 120, loss: 0.2903646230697632
step: 130, loss: 0.22843682765960693
step: 140, loss: 0.29960933327674866
step: 150, loss: 0.2813377380371094
step: 160, loss: 0.12223777174949646
step: 170, loss: 0.24516069889068604
step: 180, loss: 0.12101221829652786
step: 190, loss: 0.18587608635425568
step: 200, loss: 0.17985737323760986
step: 210, loss: 0.2819949686527252
/usr/local/lib/python3.7/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=0.44025157232704404, f1=0.482084690553746,
best_f1=0.5105633802816901
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.23139984905719757
step: 10, loss: 0.07546710968017578
step: 20, loss: 0.16790825128555298
step: 30, loss: 0.09497273713350296
step: 40, loss: 0.10462186485528946
step: 50, loss: 0.12269996106624603
step: 60, loss: 0.06910117715597153
step: 70, loss: 0.261351078748703
step: 80, loss: 0.15446820855140686
step: 90, loss: 0.11213496327400208
step: 100, loss: 0.16767434775829315
step: 110, loss: 0.15442104637622833
step: 120, loss: 0.11629240959882736
step: 130, loss: 0.21402473747730255
step: 140, loss: 0.2113310843706131
step: 150, loss: 0.312632292509079
step: 160, loss: 0.11461886763572693
step: 170, loss: 0.11366105079650879
step: 180, loss: 0.15245690941810608
step: 190, loss: 0.24931207299232483
```

```
step: 200, loss: 0.2365342080593109
step: 210, loss: 0.052091967314481735
/usr/local/lib/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=0.45348837209302323, f1=0.5095057034220533,
best_f1=0.5095057034220533
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07729168981313705
step: 10, loss: 0.05750212445855141
step: 20, loss: 0.060332439839839935
step: 30, loss: 0.15499965846538544
step: 40, loss: 0.1827387809753418
step: 50, loss: 0.10884533822536469
step: 60, loss: 0.052966780960559845
step: 70, loss: 0.16394633054733276
step: 80, loss: 0.05183327943086624
step: 90, loss: 0.15809275209903717
step: 100, loss: 0.09709998965263367
step: 110, loss: 0.015767985954880714
step: 120, loss: 0.1648351401090622
step: 130, loss: 0.08530182391405106
step: 140, loss: 0.17170460522174835
step: 150, loss: 0.07848021388053894
step: 160, loss: 0.13045963644981384
step: 170, loss: 0.18711598217487335
step: 180, loss: 0.6647425293922424
step: 190, loss: 0.10934831947088242
step: 200, loss: 0.1342773735523224
step: 210, loss: 0.18552149832248688
/usr/local/lib/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=0.45871559633027525, f1=0.5074074074074074,
best_f1=0.5074074074074074
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.041017431765794754
step: 10, loss: 0.03364713862538338
step: 20, loss: 0.0683208703994751
step: 30, loss: 0.22756369411945343
step: 40, loss: 0.04847296327352524
step: 50, loss: 0.136225163936615
step: 60, loss: 0.20988602936267853
step: 70, loss: 0.013298937119543552
step: 80, loss: 0.22682039439678192
step: 90, loss: 0.17499634623527527
step: 100, loss: 0.04699227586388588
step: 110, loss: 0.06487131863832474
step: 120, loss: 0.09065483510494232
step: 130, loss: 0.01502637006342411
step: 140, loss: 0.1022629663348198
step: 150, loss: 0.10746806114912033
step: 160, loss: 0.08899961411952972
step: 170, loss: 0.09772831201553345
step: 180, loss: 0.028799785301089287
step: 190, loss: 0.10037198662757874
step: 200, loss: 0.027637673541903496
step: 210, loss: 0.03810734674334526
/usr/local/lib/python3.7/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=0.4387351778656126, f1=0.5100401606425703,
best_f1=0.5074074074074074
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.019923627376556396
step: 10, loss: 0.04732421413064003
step: 20, loss: 0.2020084410905838
step: 30, loss: 0.01740965247154236
step: 40, loss: 0.024738095700740814
step: 50, loss: 0.12769943475723267
step: 60, loss: 0.24184638261795044
step: 70, loss: 0.10059928148984909
step: 80, loss: 0.05930298939347267
step: 90, loss: 0.05528659000992775
step: 100, loss: 0.015258423052728176
step: 110, loss: 0.07400799542665482
step: 120, loss: 0.18839779496192932
step: 130, loss: 0.030250703915953636
step: 140, loss: 0.0028400812298059464
step: 150, loss: 0.0373055599629879
step: 160, loss: 0.0778304859995842
step: 170, loss: 0.10752091556787491
```

```
step: 180, loss: 0.02145901694893837
step: 190, loss: 0.12794910371303558
step: 200, loss: 0.07017089426517487
step: 210, loss: 0.04554013907909393
/usr/local/lib/python3.7/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=0.4469820554649266, f1=0.4747967479674797,
best_f1=0.5074074074074074
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07999546080827713
step: 10, loss: 0.04684418812394142
step: 20, loss: 0.16829334199428558
step: 30, loss: 0.012941903434693813
step: 40, loss: 0.05964460223913193
step: 50, loss: 0.025577841326594353
step: 60, loss: 0.48969805240631104
step: 70, loss: 0.04412487521767616
step: 80, loss: 0.05400398001074791
step: 90, loss: 0.12031427770853043
step: 100, loss: 0.015218543820083141
step: 110, loss: 0.02179810218513012
step: 120, loss: 0.052941326051950455
step: 130, loss: 0.0694538801908493
step: 140, loss: 0.14136651158332825
step: 150, loss: 0.011539056897163391
step: 160, loss: 0.07200875133275986
step: 170, loss: 0.030682116746902466
step: 180, loss: 0.03331300616264343
step: 190, loss: 0.010649878531694412
step: 200, loss: 0.16445180773735046
step: 210, loss: 0.2081172913312912
/usr/local/lib/python3.7/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=0.43952299829642255, f1=0.48217317487266553,
best_f1=0.5074074074074074
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.043783508241176605
step: 10, loss: 0.04286842420697212
step: 20, loss: 0.06372788548469543
step: 30, loss: 0.13519102334976196
step: 40, loss: 0.10140614956617355
step: 50, loss: 0.01702398620545864
```

```
step: 60, loss: 0.07556377351284027
step: 70, loss: 0.05814630910754204
step: 80, loss: 0.004140733741223812
step: 90, loss: 0.0687933936715126
step: 100, loss: 0.02258625067770481
step: 110, loss: 0.026528626680374146
step: 120, loss: 0.17538891732692719
step: 130, loss: 0.036193620413541794
step: 140, loss: 0.07701165229082108
step: 150, loss: 0.03785306215286255
step: 160, loss: 0.002234733197838068
step: 170, loss: 0.005625082645565271
step: 180, loss: 0.020574646070599556
step: 190, loss: 0.11696479469537735
step: 200, loss: 0.07262391597032547
step: 210, loss: 0.09246574342250824
/usr/local/lib/python3.7/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=0.44086021505376344, f1=0.4695340501792115,
best_f1=0.5074074074074074
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.013369274325668812
step: 10, loss: 0.0025828881189227104
step: 20, loss: 0.013573967851698399
step: 30, loss: 0.07394220679998398
step: 40, loss: 0.07627624273300171
step: 50, loss: 0.019973194226662258
step: 60, loss: 0.015402175486087799
step: 70, loss: 0.1652706116437912
step: 80, loss: 0.002021591877564788
step: 90, loss: 0.1627814620733261
step: 100, loss: 0.002484612399712205
step: 110, loss: 0.0435514934360981
step: 120, loss: 0.0028909447137266397
step: 130, loss: 0.03395015746355057
step: 140, loss: 0.0029891813173890114
step: 150, loss: 0.07915356755256653
step: 160, loss: 0.15784451365470886
step: 170, loss: 0.048895251005887985
step: 180, loss: 0.03233252465724945
step: 190, loss: 0.2088337391614914
step: 200, loss: 0.018507840111851692
step: 210, loss: 0.06757353991270065
/usr/local/lib/python3.7/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=0.4351145038167939, f1=0.462406015037594,
best_f1=0.5074074074074074
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.015441006980836391
step: 10, loss: 0.10025814175605774
step: 20, loss: 0.036935608834028244
step: 30, loss: 0.02778504230082035
step: 40, loss: 0.03222393989562988
step: 50, loss: 0.0071099502965807915
step: 60, loss: 0.006799345836043358
step: 70, loss: 0.004366699606180191
step: 80, loss: 0.09367915242910385
step: 90, loss: 0.012889495119452477
step: 100, loss: 0.05967722088098526
step: 110, loss: 0.06242068111896515
step: 120, loss: 0.0553935207426548
step: 130, loss: 0.1410333663225174
step: 140, loss: 0.08679372817277908
step: 150, loss: 0.03911260887980461
step: 160, loss: 0.010652498342096806
step: 170, loss: 0.15101291239261627
step: 180, loss: 0.028804628178477287
step: 190, loss: 0.019888591021299362
step: 200, loss: 0.004489711485803127
step: 210, loss: 0.003124828450381756
/usr/local/lib/python3.7/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=0.4421906693711968, f1=0.4489795918367347,
best_f1=0.5074074074074074
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0041617038659751415
step: 10, loss: 0.023127729073166847
step: 20, loss: 0.012957720085978508
step: 30, loss: 0.004247849807143211
step: 40, loss: 0.2579537630081177
step: 50, loss: 0.01030450314283371
step: 60, loss: 0.04520421475172043
step: 70, loss: 0.002276407089084387
step: 80, loss: 0.02648056112229824
step: 90, loss: 0.0010403974447399378
step: 100, loss: 0.0008415335323661566
step: 110, loss: 0.010989510454237461
```

```
step: 120, loss: 0.020957518368959427
step: 130, loss: 0.0012963456101715565
step: 140, loss: 0.16297322511672974
step: 150, loss: 0.0006342598353512585
step: 160, loss: 0.003544385777786374
step: 170, loss: 0.01008114218711853
step: 180, loss: 0.0021849279291927814
step: 190, loss: 0.09030138701200485
step: 200, loss: 0.02807706408202648
step: 210, loss: 0.08903130888938904
/usr/local/lib/python3.7/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=0.42857142857142855, f1=0.4512428298279158,
best_f1=0.5074074074074074
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008711920818313956
step: 10, loss: 0.010359122417867184
step: 20, loss: 0.0016254018992185593
step: 30, loss: 0.024338314309716225
step: 40, loss: 0.01210832130163908
step: 50, loss: 0.0007629207102581859
step: 60, loss: 0.005639798939228058
step: 70, loss: 0.013860241509974003
step: 80, loss: 0.014788484200835228
step: 90, loss: 0.038377512246370316
step: 100, loss: 0.010672452859580517
step: 110, loss: 0.18804532289505005
step: 120, loss: 0.0015105202328413725
step: 130, loss: 0.0015027615008875728
step: 140, loss: 0.09815036505460739
step: 150, loss: 0.00947783887386322
step: 160, loss: 0.022610073909163475
step: 170, loss: 0.04889387637376785
step: 180, loss: 0.07689065486192703
step: 190, loss: 0.019634084776043892
step: 200, loss: 0.0054875509813427925
step: 210, loss: 0.0015259976498782635
/usr/local/lib/python3.7/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=0.41437632135306557, f1=0.4327731092436975,
best_f1=0.5074074074074074
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.005265103653073311
step: 10, loss: 0.0035461110528558493
step: 20, loss: 0.0029125092551112175
step: 30, loss: 0.14113935828208923
step: 40, loss: 0.012635786086320877
step: 50, loss: 0.02304232493042946
step: 60, loss: 0.06919068098068237
step: 70, loss: 0.03139916807413101
step: 80, loss: 0.023404141888022423
step: 90, loss: 0.13104765117168427
step: 100, loss: 0.011492365039885044
step: 110, loss: 0.005472420249134302
step: 120, loss: 0.017760783433914185
step: 130, loss: 0.0033362500835210085
step: 140, loss: 0.0027498442213982344
step: 150, loss: 0.03704897686839104
step: 160, loss: 0.0029470003210008144
step: 170, loss: 0.008316176943480968
step: 180, loss: 0.005720793269574642
step: 190, loss: 0.010891432873904705
step: 200, loss: 0.0015380759723484516
step: 210, loss: 0.016351062804460526
/usr/local/lib/python3.7/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=0.40377358490566034, f1=0.43703703703703706,
best_f1=0.5074074074074074
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02112194150686264
step: 10, loss: 0.030900241807103157
step: 20, loss: 0.0024201087653636932
step: 30, loss: 0.004489411599934101
step: 40, loss: 0.00594918429851532
step: 50, loss: 0.0029710624366998672
step: 60, loss: 0.0034407281782478094
step: 70, loss: 0.006674975622445345
step: 80, loss: 0.009175490587949753
step: 90, loss: 0.0014441125094890594
step: 100, loss: 0.012206586077809334
step: 110, loss: 0.0011382485972717404
step: 120, loss: 0.04412493109703064
step: 130, loss: 0.1062123104929924
step: 140, loss: 0.006573752965778112
step: 150, loss: 0.009343939833343029
step: 160, loss: 0.013952297158539295
step: 170, loss: 0.014428957365453243
```

step: 180, loss: 0.010949740186333656
step: 190, loss: 0.07416656613349915
step: 200, loss: 0.04339052736759186
step: 210, loss: 0.11568061262369156
/usr/local/lib/python3.7/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=0.41275797373358347, f1=0.43669724770642204,
best_f1=0.5074074074074074
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.")
2293it [00:04, 572.19it/s]
load_f1 = 0.4612736660929432
real_f1 = 0.4594127806563039
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, 378.09it/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   : 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: 0.8562628626823425
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.16457071900367737
step: 20, loss: 0.14866065979003906
step: 30, loss: 0.5040276646614075
step: 40, loss: 0.2652975022792816
step: 50, loss: 0.3099709451198578
step: 60, loss: 0.36946019530296326
step: 70, loss: 0.18170985579490662
step: 80, loss: 0.546298623085022
step: 90, loss: 0.2490377575159073
step: 100, loss: 0.22367310523986816
step: 110, loss: 0.23442304134368896
step: 120, loss: 0.4181341528892517
step: 130, loss: 0.3445049822330475
step: 140, loss: 0.34079286456108093
step: 150, loss: 0.271222859621048
step: 160, loss: 0.2237856090068817
step: 170, loss: 0.42328882217407227
step: 180, loss: 0.294037789106369
step: 190, loss: 0.14946913719177246
/usr/local/lib/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=0.4716417910447761, f1=0.4531722054380664,
best_f1=0.4531722054380664
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.32599350810050964
step: 10, loss: 0.05706140398979187
step: 20, loss: 0.0483696348965168
step: 30, loss: 0.1285291463136673
step: 40, loss: 0.496774733066655884
step: 50, loss: 0.3053276240825653
step: 60, loss: 0.13095182180404663
step: 70, loss: 0.23371034860610962
step: 80, loss: 0.13908453285694122
step: 90, loss: 0.10531173646450043
step: 100, loss: 0.12394724041223526
step: 110, loss: 0.2026384174823761
step: 120, loss: 0.4150000810623169
step: 130, loss: 0.15099549293518066
step: 140, loss: 0.25506794452667236
step: 150, loss: 0.06408865004777908
step: 160, loss: 0.018491940572857857
step: 170, loss: 0.2596144378185272
step: 180, loss: 0.2015933394432068
step: 190, loss: 0.3044692575931549
/usr/local/lib/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=0.7017543859649122, f1=0.6968838526912181,
best_f1=0.6968838526912181
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.17333191633224487
step: 10, loss: 0.1763083040714264
step: 20, loss: 0.02443438209593296
step: 30, loss: 0.03988929092884064
step: 40, loss: 0.04093324393033981
step: 50, loss: 0.11826827377080917
```

```
step: 60, loss: 0.07888365536928177
step: 70, loss: 0.2420177310705185
step: 80, loss: 0.2311423420906067
step: 90, loss: 0.06050461158156395
step: 100, loss: 0.16890867054462433
step: 110, loss: 0.45211392641067505
step: 120, loss: 0.01422809436917305
step: 130, loss: 0.08513092249631882
step: 140, loss: 0.07579810917377472
step: 150, loss: 0.15160530805587769
step: 160, loss: 0.20544281601905823
step: 170, loss: 0.04030827060341835
step: 180, loss: 0.13794070482254028
step: 190, loss: 0.12363626807928085
/usr/local/lib/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=0.7357512953367874, f1=0.7382198952879582,
best_f1=0.7382198952879582
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06042323634028435
step: 10, loss: 0.10830776393413544
step: 20, loss: 0.03939659893512726
step: 30, loss: 0.10106305778026581
step: 40, loss: 0.013145948760211468
step: 50, loss: 0.09846565872430801
step: 60, loss: 0.19319449365139008
step: 70, loss: 0.029324166476726532
step: 80, loss: 0.20521000027656555
step: 90, loss: 0.07416851818561554
step: 100, loss: 0.0656505823135376
step: 110, loss: 0.04443582892417908
step: 120, loss: 0.11376077681779861
step: 130, loss: 0.2595915198326111
step: 140, loss: 0.15336202085018158
step: 150, loss: 0.01688913255929947
step: 160, loss: 0.02333679050207138
step: 170, loss: 0.0071931700222194195
step: 180, loss: 0.16989882290363312
step: 190, loss: 0.053911611437797546
/usr/local/lib/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=0.7807486631016043, f1=0.75, best_f1=0.75
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04253234714269638
step: 10, loss: 0.054366473108530045
step: 20, loss: 0.03601953387260437
step: 30, loss: 0.05966058373451233
step: 40, loss: 0.06342075020074844
step: 50, loss: 0.059078745554395676
step: 60, loss: 0.029868675395846367
step: 70, loss: 0.03554488718509674
step: 80, loss: 0.06270436942577362
step: 90, loss: 0.008074912242591381
step: 100, loss: 0.054932694882154465
step: 110, loss: 0.06673789769411087
step: 120, loss: 0.026772111654281616
step: 130, loss: 0.025122372433543205
step: 140, loss: 0.004449798259884119
step: 150, loss: 0.03915473818778992
step: 160, loss: 0.08101249486207962
step: 170, loss: 0.054045163094997406
step: 180, loss: 0.09775663167238235
step: 190, loss: 0.1923867017030716
/usr/local/lib/python3.7/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=0.7405405405405405, f1=0.7365591397849462, best_f1=0.75
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01865379698574543
step: 10, loss: 0.0036874788347631693
step: 20, loss: 0.02162850648164749
step: 30, loss: 0.004214625805616379
step: 40, loss: 0.013069391250610352
step: 50, loss: 0.004729916341602802
step: 60, loss: 0.004658858757466078
step: 70, loss: 0.18472634255886078
step: 80, loss: 0.059995897114276886
step: 90, loss: 0.09493895620107651
step: 100, loss: 0.012729750946164131
step: 110, loss: 0.07292407751083374
```

```
step: 120, loss: 0.057513874024152756
step: 130, loss: 0.01566443033516407
step: 140, loss: 0.012511517852544785
step: 150, loss: 0.0881476029753685
step: 160, loss: 0.03452416509389877
step: 170, loss: 0.09766115993261337
step: 180, loss: 0.006861855275928974
step: 190, loss: 0.003892763750627637
/usr/local/lib/python3.7/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=0.7520891364902507, f1=0.7493112947658401, best_f1=0.75
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006578109227120876
step: 10, loss: 0.052267204970121384
step: 20, loss: 0.0010538690257817507
step: 30, loss: 0.03197978436946869
step: 40, loss: 0.03788352012634277
step: 50, loss: 0.02734304592013359
step: 60, loss: 0.009417633526027203
step: 70, loss: 0.003349879989400506
step: 80, loss: 0.13091015815734863
step: 90, loss: 0.0030895129311829805
step: 100, loss: 0.005505729466676712
step: 110, loss: 0.019435012713074684
step: 120, loss: 0.02476588450372219
step: 130, loss: 0.0032069445587694645
step: 140, loss: 0.018204577267169952
step: 150, loss: 0.005548818502575159
step: 160, loss: 0.002483718330040574
step: 170, loss: 0.0028426055796444416
step: 180, loss: 0.006610897369682789
step: 190, loss: 0.04067108780145645
/usr/local/lib/python3.7/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=0.7559055118110236, f1=0.7195767195767196, best_f1=0.75
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0014058771776035428
step: 10, loss: 0.010480690747499466
step: 20, loss: 0.004777503665536642
step: 30, loss: 0.003002490848302841
step: 40, loss: 0.15334944427013397
step: 50, loss: 0.010723496787250042
```

```
step: 60, loss: 0.037877634167671204
step: 70, loss: 0.0033447009045630693
step: 80, loss: 0.004825165960937738
step: 90, loss: 0.0018207021057605743
step: 100, loss: 0.20287667214870453
step: 110, loss: 0.007783415727317333
step: 120, loss: 0.08014723658561707
step: 130, loss: 0.0031975216697901487
step: 140, loss: 0.010950236581265926
step: 150, loss: 0.06490515917539597
step: 160, loss: 0.01912875846028328
step: 170, loss: 0.01363363116979599
step: 180, loss: 0.023227276280522346
step: 190, loss: 0.0431981086730957
/usr/local/lib/python3.7/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=0.7455012853470436, f1=0.743142144638404, best_f1=0.75
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2172415405511856
step: 10, loss: 0.0034309346228837967
step: 20, loss: 0.14433152973651886
step: 30, loss: 0.06333014369010925
step: 40, loss: 0.00883129145950079
step: 50, loss: 0.0029986875597387552
step: 60, loss: 0.004717483185231686
step: 70, loss: 0.0004816264845430851
step: 80, loss: 0.011903837323188782
step: 90, loss: 0.020100325345993042
step: 100, loss: 0.0014285410288721323
step: 110, loss: 0.002583723049610853
step: 120, loss: 0.0674452930688858
step: 130, loss: 0.007769250776618719
step: 140, loss: 0.0020829858258366585
step: 150, loss: 0.005577942822128534
step: 160, loss: 0.07913389801979065
step: 170, loss: 0.06691926717758179
step: 180, loss: 0.07511626183986664
step: 190, loss: 0.00816453155130148
/usr/local/lib/python3.7/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=0.7493403693931397, f1=0.7168831168831169, best_f1=0.75
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.0014377835905179381
step: 10, loss: 0.001997358165681362
step: 20, loss: 0.007418425288051367
step: 30, loss: 0.008117283694446087
step: 40, loss: 0.0012784932041540742
step: 50, loss: 0.0066302623599767685
step: 60, loss: 0.1507616937160492
step: 70, loss: 0.007806878536939621
step: 80, loss: 0.0016922628274187446
step: 90, loss: 0.004298159386962652
step: 100, loss: 0.017009085044264793
step: 110, loss: 0.024779127910733223
step: 120, loss: 0.00242599262855947
step: 130, loss: 0.0012166649103164673
step: 140, loss: 0.005118592642247677
step: 150, loss: 0.0015548779629170895
step: 160, loss: 0.0008071210468187928
step: 170, loss: 0.001264885999262333
step: 180, loss: 0.009494726546108723
step: 190, loss: 0.00920331571251154
/usr/local/lib/python3.7/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=0.7345844504021448, f1=0.7282321899736147, best_f1=0.75
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0026649204082787037
step: 10, loss: 0.011290152557194233
step: 20, loss: 0.0020829602144658566
step: 30, loss: 0.0022144028916954994
step: 40, loss: 0.0028773811645805836
step: 50, loss: 0.001332474290393293
step: 60, loss: 0.0043749697506427765
step: 70, loss: 0.0008162546437233686
step: 80, loss: 0.0046321433037519455
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 90, loss: 0.0013402156764641404
step: 100, loss: 0.0011141253635287285
step: 110, loss: 0.0014098434476181865
step: 120, loss: 0.001461922307498753
step: 130, loss: 0.0021245493553578854
step: 140, loss: 0.0006588895339518785
step: 150, loss: 0.006615896709263325
step: 160, loss: 0.001122821937315166
step: 170, loss: 0.058531925082206726
step: 180, loss: 0.04201061651110649
step: 190, loss: 0.005394786596298218
```

```
/usr/local/lib/python3.7/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=0.7567567567567567, f1=0.7049180327868853, best_f1=0.75
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0016755149699747562
step: 10, loss: 0.005513709504157305
step: 20, loss: 0.003835184033960104
step: 30, loss: 0.07912891358137131
step: 40, loss: 0.008966632187366486
step: 50, loss: 0.0014443728141486645
step: 60, loss: 0.00046374715748243034
step: 70, loss: 0.0021270031575113535
step: 80, loss: 0.0016072371508926153
step: 90, loss: 0.20068852603435516
step: 100, loss: 0.003975106868892908
step: 110, loss: 0.008296751417219639
step: 120, loss: 0.001254492555744946
step: 130, loss: 0.005653807893395424
step: 140, loss: 0.02282690815627575
step: 150, loss: 0.0023059339728206396
step: 160, loss: 0.017518723383545876
step: 170, loss: 0.0056769223883748055
step: 180, loss: 0.0014648601645603776
step: 190, loss: 0.1536587029695511
/usr/local/lib/python3.7/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=0.7317073170731707, f1=0.7154471544715448, best_f1=0.75
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008345250971615314
step: 10, loss: 0.0032533679623156786
step: 20, loss: 0.00221968162804842
step: 30, loss: 0.004009414929896593
step: 40, loss: 0.0010042381472885609
step: 50, loss: 0.011640372686088085
step: 60, loss: 0.002024967223405838
step: 70, loss: 0.0005866261781193316
step: 80, loss: 0.0038511166349053383
step: 90, loss: 0.041880447417497635
step: 100, loss: 0.0052299718372523785
step: 110, loss: 0.0014124992303550243
step: 120, loss: 0.0021639158949255943
step: 130, loss: 0.052662163972854614
```

```
step: 140, loss: 0.004107967484742403
step: 150, loss: 0.018198223784565926
step: 160, loss: 0.007957279682159424
step: 170, loss: 0.001885897945612669
step: 180, loss: 0.00280465604737401
step: 190, loss: 0.02137838862836361
/usr/local/lib/python3.7/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=0.7317073170731707, f1=0.7162534435261707, best_f1=0.75
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.12202328443527222
step: 10, loss: 0.002188280690461397
step: 20, loss: 0.0027806032449007034
step: 30, loss: 0.007026742212474346
step: 40, loss: 0.0009155264706350863
step: 50, loss: 0.0011608425993472338
step: 60, loss: 0.0035965361166745424
step: 70, loss: 0.00914947409182787
step: 80, loss: 0.001285324222408235
step: 90, loss: 0.007822778075933456
step: 100, loss: 0.014757073484361172
step: 110, loss: 0.0005374421016313136
step: 120, loss: 0.0014354687882587314
step: 130, loss: 0.0461464524269104
step: 140, loss: 0.0027879599947482347
step: 150, loss: 0.0008925607544369996
step: 160, loss: 0.005988023243844509
step: 170, loss: 0.21258771419525146
step: 180, loss: 0.0016485777450725436
step: 190, loss: 0.0007395753054879606
/usr/local/lib/python3.7/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=0.7272727272727273, f1=0.7138964577656676, best_f1=0.75
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0011328815016895533
step: 10, loss: 0.0023356920573860407
step: 20, loss: 0.0005165491602383554
step: 30, loss: 0.0012348806485533714
step: 40, loss: 0.0008135930984281003
step: 50, loss: 0.001165401772595942
step: 60, loss: 0.0006674696342088282
step: 70, loss: 0.0018377716187387705
```

```
step: 80, loss: 0.0405329130589962
step: 90, loss: 0.0545785091817379
step: 100, loss: 0.0005807693814858794
step: 110, loss: 0.006647940259426832
step: 120, loss: 0.0015899526188150048
step: 130, loss: 0.004809225909411907
step: 140, loss: 0.005679886322468519
step: 150, loss: 0.001350156613625586
step: 160, loss: 0.008555784821510315
step: 170, loss: 0.00040892360266298056
step: 180, loss: 0.001959078246727586
step: 190, loss: 0.0009524824563413858
/usr/local/lib/python3.7/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=0.7308781869688386, f1=0.6961651917404129, best_f1=0.75
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.")
2049it [00:06, 316.13it/s]
load_f1 = 0.7382920110192837
```

```
real_f1 = 0.745308310991957
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, 362.62it/s]
```

4.2 DITTO TEXTUAL

4.2.1 Abt-Buy - Running the matcher

```python
# 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 O2:  FP16 training with FP32 batchnorm and FP32
master weights.
```

417

```
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: 0.846549928188324
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.21384912729263306
step: 20, loss: 0.14727512001991272
step: 30, loss: 0.24446696043014526
step: 40, loss: 0.3136391043663025
step: 50, loss: 0.37785282731056213
step: 60, loss: 0.43352562189102173
step: 70, loss: 0.308724582195282
step: 80, loss: 0.2513115108013153
step: 90, loss: 0.4104374349117279
step: 100, loss: 0.2350817173719406
step: 110, loss: 0.18789924681186676
step: 120, loss: 0.5299362540245056
```

```
step: 130, loss: 0.4188780188560486
step: 140, loss: 0.47661417722702026
step: 150, loss: 0.09693492949008942
step: 160, loss: 0.3466000258922577
step: 170, loss: 0.24337105453014374
/usr/local/lib/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=0.2907268170426065, f1=0.24646983311938386,
best_f1=0.24646983311938386
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.42361608147621155
step: 10, loss: 0.24290873110294342
step: 20, loss: 0.3396557569503784
step: 30, loss: 0.2972502112388611
step: 40, loss: 0.21070623397827148
step: 50, loss: 0.2896919250488281
step: 60, loss: 0.18321597576141357
step: 70, loss: 0.23228619992733002
step: 80, loss: 0.09910836815834045
step: 90, loss: 0.2618633210659027
step: 100, loss: 0.11349281668663025
step: 110, loss: 0.30151864886283875
step: 120, loss: 0.09044674783945084
step: 130, loss: 0.05945061892271042
step: 140, loss: 0.2074000984430313
step: 150, loss: 0.15411125123500824
step: 160, loss: 0.18712730705738068
step: 170, loss: 0.1089579164981842
/usr/local/lib/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=0.6477024070021883, f1=0.6233183856502242,
best_f1=0.6233183856502242
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.157114639878273
```

```
step: 10, loss: 0.12349268794059753
step: 20, loss: 0.04277724772691727
step: 30, loss: 0.38413605093955994
step: 40, loss: 0.011858168989419937
step: 50, loss: 0.18130715191364288
step: 60, loss: 0.2608896493911743
step: 70, loss: 0.2945590913295746
step: 80, loss: 0.16900953650474548
step: 90, loss: 0.23851680755615234
step: 100, loss: 0.03689416125416756
step: 110, loss: 0.13638626039028168
step: 120, loss: 0.028339054435491562
step: 130, loss: 0.17442795634269714
step: 140, loss: 0.015921350568532944
step: 150, loss: 0.12858940660953522
step: 160, loss: 0.07268083095550537
step: 170, loss: 0.22855377197265625
/usr/local/lib/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=0.7405405405405405, f1=0.6984126984126984,
best_f1=0.6984126984126984
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08215069025754929
step: 10, loss: 0.08592269569635391
step: 20, loss: 0.02895314246416092
step: 30, loss: 0.17120403051376343
step: 40, loss: 0.099124975502491
step: 50, loss: 0.05228649824857712
step: 60, loss: 0.16615648567676544
step: 70, loss: 0.010889961384236813
step: 80, loss: 0.027010781690478325
step: 90, loss: 0.06087670102715492
step: 100, loss: 0.06346023082733154
step: 110, loss: 0.14496682584285736
step: 120, loss: 0.17454279959201813
step: 130, loss: 0.06603571772575378
step: 140, loss: 0.030099380761384964
step: 150, loss: 0.05142296105623245
step: 160, loss: 0.09828471392393112
step: 170, loss: 0.04875842481851578
/usr/local/lib/python3.7/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=0.7240506329113925, f1=0.7002518891687658,
best_f1=0.6984126984126984
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07941460609436035
step: 10, loss: 0.1082068607211113
step: 20, loss: 0.03650841489434242
step: 30, loss: 0.16683518886566162
step: 40, loss: 0.021395744755864143
step: 50, loss: 0.025644950568675995
step: 60, loss: 0.004081200808286667
step: 70, loss: 0.08487547934055328
step: 80, loss: 0.014791733585298061
step: 90, loss: 0.0921589657664299
step: 100, loss: 0.03885551169514656
step: 110, loss: 0.1373131424188614
step: 120, loss: 0.13631895184516907
step: 130, loss: 0.029522234573960304
step: 140, loss: 0.03829227760434151
step: 150, loss: 0.048949360847473145
step: 160, loss: 0.05331040546298027
step: 170, loss: 0.048667244613170624
/usr/local/lib/python3.7/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=0.7286821705426356, f1=0.685, best_f1=0.6984126984126984
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02714608982205391
step: 10, loss: 0.09025851637125015
step: 20, loss: 0.035152241587638855
step: 30, loss: 0.053470585495233536
step: 40, loss: 0.013123015873134136
step: 50, loss: 0.04003998264670372
step: 60, loss: 0.022680656984448433
step: 70, loss: 0.004066824913024902
step: 80, loss: 0.20483875274658203
step: 90, loss: 0.04535358399152756
step: 100, loss: 0.20315150916576385
step: 110, loss: 0.11110392212867737
step: 120, loss: 0.1692485213279724
step: 130, loss: 0.3056694567203522
step: 140, loss: 0.009307385422289371
step: 150, loss: 0.3852970600128174
```

```
step: 160, loss: 0.04916002228856087
step: 170, loss: 0.015759317204356194
/usr/local/lib/python3.7/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=0.7268041237113402, f1=0.6813725490196079,
best_f1=0.6984126984126984
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02127944491803646
step: 10, loss: 0.00475477147847414
step: 20, loss: 0.0198493842035532
step: 30, loss: 0.10608416050672531
step: 40, loss: 0.0148554015904665
step: 50, loss: 0.029195230454206467
step: 60, loss: 0.25656774640083313
step: 70, loss: 0.00846855714917183
step: 80, loss: 0.029064955189824104
step: 90, loss: 0.033885419368743896
step: 100, loss: 0.005566906183958054
step: 110, loss: 0.011969661340117455
step: 120, loss: 0.21391771733760834
step: 130, loss: 0.1469971239566803
step: 140, loss: 0.010301053524017334
step: 150, loss: 0.04446905478835106
step: 160, loss: 0.013664853759109974
step: 170, loss: 0.1785700023174286
/usr/local/lib/python3.7/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=0.7180851063829787, f1=0.6898395721925135,
best_f1=0.6984126984126984
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.059110838919878006
step: 10, loss: 0.014437716454267502
step: 20, loss: 0.05101816728711128
step: 30, loss: 0.06109825149178505
step: 40, loss: 0.021749090403318405
step: 50, loss: 0.11845327913761139
step: 60, loss: 0.023875916376709938
step: 70, loss: 0.057708170264959335
step: 80, loss: 0.004061523359268904
step: 90, loss: 0.1057099997997284
step: 100, loss: 0.013519179075956345
step: 110, loss: 0.0466693714261055
```

```
step: 120, loss: 0.01641143672168255
step: 130, loss: 0.0051255906000733376
step: 140, loss: 0.005804760381579399
step: 150, loss: 0.15102745592594147
step: 160, loss: 0.02776653878390789
step: 170, loss: 0.01543737854808569
/usr/local/lib/python3.7/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=0.7212276214833758, f1=0.6962962962962963,
best_f1=0.6984126984126984
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00475778104737401
step: 10, loss: 0.10249608755111694
step: 20, loss: 0.13305748999118805
step: 30, loss: 0.1017623022198677
step: 40, loss: 0.14053623378276825
step: 50, loss: 0.08125433325767517
step: 60, loss: 0.008577299304306507
step: 70, loss: 0.012360382825136185
step: 80, loss: 0.057217735797166824
step: 90, loss: 0.023509949445724487
step: 100, loss: 0.049436263740062714
step: 110, loss: 0.06677766889333725
step: 120, loss: 0.01824929006397724
step: 130, loss: 0.0018922097515314817
step: 140, loss: 0.005495480261743069
step: 150, loss: 0.037767067551612854
step: 160, loss: 0.02379184402525425
step: 170, loss: 0.03176990896463394
/usr/local/lib/python3.7/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=0.7262872628726287, f1=0.6986666666666667,
best_f1=0.6984126984126984
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.024874428287148476
step: 10, loss: 0.011784031987190247
step: 20, loss: 0.004495501518249512
step: 30, loss: 0.03061024472117424
step: 40, loss: 0.003593273228034377
step: 50, loss: 0.006767392158508301
step: 60, loss: 0.025543304160237312
step: 70, loss: 0.004275710321962833
```

```
step: 80, loss: 0.008730929344892502
step: 90, loss: 0.011370300315320492
step: 100, loss: 0.057431790977716446
step: 110, loss: 0.0007117881905287504
step: 120, loss: 0.026705851778388023
step: 130, loss: 0.1020040437579155
step: 140, loss: 0.029118737205863
step: 150, loss: 0.0016576729249209166
step: 160, loss: 0.06936217099428177
step: 170, loss: 0.017659470438957214
/usr/local/lib/python3.7/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=0.7394957983193278, f1=0.6956521739130433,
best_f1=0.6984126984126984
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008569757919758558
step: 10, loss: 0.026814335957169533
step: 20, loss: 0.01574811153113842
step: 30, loss: 0.005117741413414478
step: 40, loss: 0.21816898882389069
step: 50, loss: 0.0036729092244058847
step: 60, loss: 0.00884929671883583
step: 70, loss: 0.007854265160858631
step: 80, loss: 0.06539975851774216
step: 90, loss: 0.019663441926240092
step: 100, loss: 0.0016566303092986345
step: 110, loss: 0.02991321310400963
step: 120, loss: 0.08633332699537277
step: 130, loss: 0.008159532211720943
step: 140, loss: 0.07921252399682999
step: 150, loss: 0.0035327638033777475
step: 160, loss: 0.08930165320634842
step: 170, loss: 0.05046509951353073
/usr/local/lib/python3.7/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=0.7150259067357513, f1=0.6564885496183206,
best_f1=0.6984126984126984
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010847226716578007
step: 10, loss: 0.011227929033339024
step: 20, loss: 0.019241902977228165
step: 30, loss: 0.004051906988024712
```

```
step: 40, loss: 0.004927445203065872
step: 50, loss: 0.016715966165065765
step: 60, loss: 0.026588989421725273
step: 70, loss: 0.008946651592850685
step: 80, loss: 0.12215344607830048
step: 90, loss: 0.0018732595490291715
step: 100, loss: 0.07627738267183304
step: 110, loss: 0.02507144771516323
step: 120, loss: 0.02543114870786667
step: 130, loss: 0.03081163763999939
step: 140, loss: 0.005801243707537651
step: 150, loss: 0.01694730669260025
step: 160, loss: 0.004458484705537558
step: 170, loss: 0.025764059275388718
/usr/local/lib/python3.7/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=0.7139240506329114, f1=0.6813725490196079,
best_f1=0.6984126984126984
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003546877996996045
step: 10, loss: 0.06300707161426544
step: 20, loss: 0.0014050058089196682
step: 30, loss: 0.0013903815997764468
step: 40, loss: 0.003102516988923365
step: 50, loss: 0.0017493950435891747
step: 60, loss: 0.0006721844547428191
step: 70, loss: 0.004898956045508385
step: 80, loss: 0.0014340280322358012
step: 90, loss: 0.0008613343816250563
step: 100, loss: 0.0009778960375115275
step: 110, loss: 0.023336950689554214
step: 120, loss: 0.007077925838530064
step: 130, loss: 0.006416539195924997
step: 140, loss: 0.0017410935834050179
step: 150, loss: 0.0031908140517771244
step: 160, loss: 0.0012026039185002446
step: 170, loss: 0.0041062175296247005
/usr/local/lib/python3.7/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=0.7166666666666667, f1=0.6927374301675977,
best_f1=0.6984126984126984
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
```

```
step: 0, loss: 0.001968675060197711
step: 10, loss: 0.002344615524634719
step: 20, loss: 0.011279827915132046
step: 30, loss: 0.033747054636478424
step: 40, loss: 0.00090132059995085
step: 50, loss: 0.21145674586296082
step: 60, loss: 0.002648168709129095
step: 70, loss: 0.0021362826228141785
step: 80, loss: 0.005708268843591213
step: 90, loss: 0.04799745976924896
step: 100, loss: 0.006653282791376114
step: 110, loss: 0.0036332898307591677
step: 120, loss: 0.004287000745534897
step: 130, loss: 0.0027089843060821295
step: 140, loss: 0.00543819647282362
step: 150, loss: 0.03182568401098251
step: 160, loss: 0.05098462849855423
step: 170, loss: 0.0008320931810885668
/usr/local/lib/python3.7/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=0.7247956403269755, f1=0.6991869918699187,
best_f1=0.6984126984126984
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005451960489153862
step: 10, loss: 0.002433191053569317
step: 20, loss: 0.020416809245944023
step: 30, loss: 0.04189841076731682
step: 40, loss: 0.0005028037703596056
step: 50, loss: 0.002937138546258211
step: 60, loss: 0.0007449996192008257
step: 70, loss: 0.005136698484420776
step: 80, loss: 0.0012278524227440357
step: 90, loss: 0.0036347422283142805
step: 100, loss: 0.004637017846107483
step: 110, loss: 0.0056334906257689
step: 120, loss: 0.020624201744794846
step: 130, loss: 0.0012358510866761208
step: 140, loss: 0.0013828924857079983
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 150, loss: 0.04521197825670242
step: 160, loss: 0.0090520940721035
step: 170, loss: 0.0017555560916662216
/usr/local/lib/python3.7/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=0.7237569060773481, f1=0.7036011080332409,
best_f1=0.6984126984126984
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, 438.05it/s]
load_f1 = 0.7086614173228347
real_f1 = 0.7015706806282723
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, 399.39it/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 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: 0.7981746196746826
step: 10, loss: 0.47928956151008606
step: 20, loss: 0.5966882705688477
step: 30, loss: 0.4985109567642212
step: 40, loss: 0.3850580155849457
step: 50, loss: 0.3307805359363556
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 60, loss: 0.16282522678375244
step: 70, loss: 0.23142722249031067
step: 80, loss: 0.24285255372524261
step: 90, loss: 0.13599742949008942
step: 100, loss: 0.12443788349628448
step: 110, loss: 0.11283382773399353
step: 120, loss: 0.05452249199151993
step: 130, loss: 0.028280459344387054
step: 140, loss: 0.09828140586614609
step: 150, loss: 0.11178809404373169
step: 160, loss: 0.15617595613002777
step: 170, loss: 0.013633045367896557
step: 180, loss: 0.014914633706212044
step: 190, loss: 0.2601086497306824
```

```
step: 200, loss: 0.13426411151885986
step: 210, loss: 0.04637556150555611
step: 220, loss: 0.015230854973196983
step: 230, loss: 0.04160868003964424
/usr/local/lib/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=0.9317180616740088, f1=0.9318435754189943,
best_f1=0.9318435754189943
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1187271773815155
step: 10, loss: 0.11351843923330307
step: 20, loss: 0.0360351987183094
step: 30, loss: 0.008694672025740147
step: 40, loss: 0.08736088871955872
step: 50, loss: 0.04958823695778847
step: 60, loss: 0.04659897834062576
step: 70, loss: 0.024463798850774765
step: 80, loss: 0.01974346674978733
step: 90, loss: 0.016679903492331505
step: 100, loss: 0.18471936881542206
step: 110, loss: 0.037987180054187775
step: 120, loss: 0.06907433271408081
step: 130, loss: 0.08151856064796448
step: 140, loss: 0.15185606479644775
step: 150, loss: 0.11089866608381271
step: 160, loss: 0.06191747635602951
step: 170, loss: 0.043284520506858826
step: 180, loss: 0.05666545405983925
step: 190, loss: 0.1227840781211853
step: 200, loss: 0.04269307479262352
step: 210, loss: 0.09872410446405411
step: 220, loss: 0.004611428827047348
step: 230, loss: 0.017230499535799026
/usr/local/lib/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=0.9481808158765159, f1=0.9358830146231721,
```

```
best_f1=0.9358830146231721
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06462721526622772
step: 10, loss: 0.03604961931705475
step: 20, loss: 0.005403400398790836
step: 30, loss: 0.053473781794309616
step: 40, loss: 0.05887873098254204
step: 50, loss: 0.03519086539745331
step: 60, loss: 0.12665115296840668
step: 70, loss: 0.031079763546586037
step: 80, loss: 0.034283678978681564
step: 90, loss: 0.04823361337184906
step: 100, loss: 0.0052452655509114265
step: 110, loss: 0.1020960584282875
step: 120, loss: 0.01358024962246418
step: 130, loss: 0.001587537582963705
step: 140, loss: 0.0026156583335250616
step: 150, loss: 0.01029429491609335
step: 160, loss: 0.009446787647902966
step: 170, loss: 0.13460826873779297
step: 180, loss: 0.007128271274268627
step: 190, loss: 0.013460456393659115
step: 200, loss: 0.0076395366340875626
step: 210, loss: 0.009566099382936954
step: 220, loss: 0.018931657075881958
step: 230, loss: 0.13945169746875763
/usr/local/lib/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=0.9491150442477876, f1=0.9461883408071748,
best_f1=0.9461883408071748
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007177019026130438
step: 10, loss: 0.00611441396176815
step: 20, loss: 0.05844248831272125
step: 30, loss: 0.019985785707831383
step: 40, loss: 0.026715505868196487
step: 50, loss: 0.007457279600203037
step: 60, loss: 0.001395341125316918
step: 70, loss: 0.014799901284277439
```

```
step: 80, loss: 0.12532952427864075
step: 90, loss: 0.005480160471051931
step: 100, loss: 0.0028589756693691015
step: 110, loss: 0.0257476344704628
step: 120, loss: 0.005595740862190723
step: 130, loss: 0.0021560548339039087
step: 140, loss: 0.006380511447787285
step: 150, loss: 0.0049771564081311123
step: 160, loss: 0.008680887520313263
step: 170, loss: 0.1059633120894432
step: 180, loss: 0.0047697811387479305
step: 190, loss: 0.031264662742614746
step: 200, loss: 0.004618074744939804
step: 210, loss: 0.07072529941797256
step: 220, loss: 0.016803890466690063
step: 230, loss: 0.00422531645745039
/usr/local/lib/python3.7/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=0.9387755102040816, f1=0.9345579793340988,
best_f1=0.9461883408071748
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009079794399440289
step: 10, loss: 0.007084952667355537
step: 20, loss: 0.003943896386772394
step: 30, loss: 0.0006764406571164727
step: 40, loss: 0.004474794492125511
step: 50, loss: 0.004131501540541649
step: 60, loss: 0.07746041566133499
step: 70, loss: 0.0038793806452304125
step: 80, loss: 0.002897412283346057
step: 90, loss: 0.0076660169288516045
step: 100, loss: 0.04943598434329033
step: 110, loss: 0.004660635720938444
step: 120, loss: 0.02186436578631401
step: 130, loss: 0.09700558334589005
step: 140, loss: 0.003687991062179208
step: 150, loss: 0.004335220903158188
step: 160, loss: 0.03458915278315544
step: 170, loss: 0.17893041670322418
step: 180, loss: 0.002665470587089658
step: 190, loss: 0.003291034372523427
step: 200, loss: 0.018258055672049522
step: 210, loss: 0.0027353139594197273
step: 220, loss: 0.0011099124094471335
step: 230, loss: 0.04938459396362305
```

```
/usr/local/lib/python3.7/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=0.9486607142857143, f1=0.9415730337078653,
best_f1=0.9461883408071748
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0037016270216554403
step: 10, loss: 0.02308814972639084
step: 20, loss: 0.001052799983881414
step: 30, loss: 0.003159956308081746
step: 40, loss: 0.04059892147779465
step: 50, loss: 0.09616944193840027
step: 60, loss: 0.010274147614836693
step: 70, loss: 0.0011672945693135262
step: 80, loss: 0.0023021858651190996
step: 90, loss: 0.0014660876477137208
step: 100, loss: 0.0037294665817171335
step: 110, loss: 0.2289830595254898
step: 120, loss: 0.008185871876776218
step: 130, loss: 0.0038335449062287807
step: 140, loss: 0.00817815400660038
step: 150, loss: 0.004081493243575096
step: 160, loss: 0.0014121417189016938
step: 170, loss: 0.002939034253358841
step: 180, loss: 0.002280141692608595
step: 190, loss: 0.030609775334596634
step: 200, loss: 0.01247985940426588
step: 210, loss: 0.001266294508241117
step: 220, loss: 0.0005463407724164426
step: 230, loss: 0.0020001695957034826
/usr/local/lib/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=0.9512195121951219, f1=0.9498327759197325,
best_f1=0.9498327759197325
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.18865913152694702
step: 10, loss: 0.0009666946716606617
step: 20, loss: 0.00300309294834733
step: 30, loss: 0.01454308070242405
```

```
step: 40, loss: 0.002978892996907234
step: 50, loss: 0.0007001270423643291
step: 60, loss: 0.04247775301337242
step: 70, loss: 0.008524863049387932
step: 80, loss: 0.011748308315873146
step: 90, loss: 0.005750618875026703
step: 100, loss: 0.0006076113204471767
step: 110, loss: 0.007204617839306593
step: 120, loss: 0.013148884288966656
step: 130, loss: 0.019146496430039406
step: 140, loss: 0.0015004774322733283
step: 150, loss: 0.003621876472607255
step: 160, loss: 0.04694371670484543
step: 170, loss: 0.00034149823477491736
step: 180, loss: 0.0007529232534579933
step: 190, loss: 0.006398649420589209
step: 200, loss: 0.0061067561618983746
step: 210, loss: 0.0006421646103262901
step: 220, loss: 0.003520811442285776
step: 230, loss: 0.08416302502155304
/usr/local/lib/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=0.9518477043673013, f1=0.9435665914221218,
best_f1=0.9435665914221218
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0011834357865154743
step: 10, loss: 0.34571516513824463
step: 20, loss: 0.0007321450975723565
step: 30, loss: 0.0010351238306611776
step: 40, loss: 0.0009783478453755379
step: 50, loss: 0.0018574680434539914
step: 60, loss: 0.00046295669744722545
step: 70, loss: 0.003315561218187213
step: 80, loss: 0.004633861593902111
step: 90, loss: 0.003979984670877457
step: 100, loss: 0.0005346803227439523
step: 110, loss: 0.0020469047594815493
step: 120, loss: 0.006996171083301306
step: 130, loss: 0.004177777096629143
step: 140, loss: 0.0007914719171822071
step: 150, loss: 0.0012727963039651513
```

```
step: 160, loss: 0.003525302279740572
step: 170, loss: 0.0015382288256660104
step: 180, loss: 0.011366945691406727
step: 190, loss: 0.0009278489160351455
step: 200, loss: 0.014774810522794724
step: 210, loss: 0.004021852742880583
step: 220, loss: 0.0004040678031742573
step: 230, loss: 0.20406535267829895
/usr/local/lib/python3.7/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=0.9518477043673013, f1=0.9410430839002268,
best_f1=0.9435665914221218
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002408822299912572
step: 10, loss: 0.0017490088939666748
step: 20, loss: 0.004784193821251392
step: 30, loss: 0.0015192677965387702
step: 40, loss: 0.05053116753697395
step: 50, loss: 0.0005720055196434259
step: 60, loss: 0.008777856826782227
step: 70, loss: 0.09685872495174408
step: 80, loss: 0.075691498816011334
step: 90, loss: 0.0016874163411557674
step: 100, loss: 0.000855709717143327
step: 110, loss: 0.0003974508144892752
step: 120, loss: 0.05079597234725952
step: 130, loss: 0.0016560337971895933
step: 140, loss: 0.04256248474121094
step: 150, loss: 0.001489349640905857
step: 160, loss: 0.0006784720462746918
step: 170, loss: 0.00019556567713152617
step: 180, loss: 0.0017376340692862868
step: 190, loss: 0.004045199602842331
step: 200, loss: 0.00021754980843979865
step: 210, loss: 0.0003640512877609581
step: 220, loss: 0.018482958897948265
step: 230, loss: 0.002803141251206398
/usr/local/lib/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=0.9565217391304347, f1=0.9451287793952967,
```

```
best_f1=0.9451287793952967
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002505892189219594
step: 10, loss: 0.0001385996729368344
step: 20, loss: 0.0002799782669171691
step: 30, loss: 0.0004420682380441576
step: 40, loss: 0.002517684595659375
step: 50, loss: 0.031951893121004105
step: 60, loss: 0.0416036881506443
step: 70, loss: 0.00406602630391717
step: 80, loss: 0.0011823491659015417
step: 90, loss: 0.00018366152653470635
step: 100, loss: 0.00041908628190867603
step: 110, loss: 0.00017749816470313817
step: 120, loss: 0.01474534161388874
step: 130, loss: 0.009812125004827976
step: 140, loss: 0.05412349849939346
step: 150, loss: 0.0004374094132799655
step: 160, loss: 0.0008944948785938323
step: 170, loss: 0.0010266833705827594
step: 180, loss: 0.0070130350068211555
step: 190, loss: 0.0007549207657575607
step: 200, loss: 0.003374811029061675
step: 210, loss: 0.0004711260844487697
step: 220, loss: 0.002592662116512656
step: 230, loss: 0.021909721195697784
/usr/local/lib/python3.7/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=0.9530201342281879, f1=0.9503386004514672,
best_f1=0.9451287793952967
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006287691066972911
step: 10, loss: 0.0007832198170945048
step: 20, loss: 0.0002618947473820299
step: 30, loss: 0.00017239000590052456
step: 40, loss: 0.0028181227389723063
step: 50, loss: 0.0006304644048213959
step: 60, loss: 0.014190525747835636
step: 70, loss: 0.0018445922760292888
step: 80, loss: 0.004529297351837158
step: 90, loss: 0.0063543482683598995
step: 100, loss: 0.000562359462492168
step: 110, loss: 0.0002390672598266974
```

```
step: 120, loss: 0.0016845723148435354
step: 130, loss: 0.0002416400529909879
step: 140, loss: 0.00018677336629480124
step: 150, loss: 0.0002602917666081339
step: 160, loss: 0.0008026125724427402
step: 170, loss: 0.00018674526654649526
step: 180, loss: 0.0011836475459858775
step: 190, loss: 0.004612443502992392
step: 200, loss: 0.0005909692845307291
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 210, loss: 0.0001641264243517071
step: 220, loss: 0.0003336627851240337
step: 230, loss: 0.009352847933769226
/usr/local/lib/python3.7/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=0.9501661129568106, f1=0.9518477043673013,
best_f1=0.9451287793952967
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003191999567206949
step: 10, loss: 0.0002891384647227824
step: 20, loss: 0.00024258316261693835
step: 30, loss: 0.0421210341155529
step: 40, loss: 0.0046315984800457954
step: 50, loss: 0.0001861527271103114
step: 60, loss: 0.009885805658996105
step: 70, loss: 0.0002621510357130319
step: 80, loss: 0.00015110449749045074
step: 90, loss: 0.0004283634480088949
step: 100, loss: 0.00026664554025046530
step: 110, loss: 0.0009617353207431734
step: 120, loss: 0.00028355655376799405
step: 130, loss: 0.0018560473108664155
step: 140, loss: 0.0001273863308597356
step: 150, loss: 7.586924766656011e-05
step: 160, loss: 0.00012133269046898931
step: 170, loss: 0.000478225527331233
step: 180, loss: 8.12937505543232e-05
step: 190, loss: 0.0007234029471874237
step: 200, loss: 0.0006055555422790349
step: 210, loss: 0.02047288790345192
step: 220, loss: 0.007192941382527351
step: 230, loss: 0.00016017093730624765
/usr/local/lib/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=0.9588431590656283, f1=0.9506726457399103,
best_f1=0.9506726457399103
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0012215584283694625
step: 10, loss: 0.00012022460578009486
step: 20, loss: 0.0002973934751935303
step: 30, loss: 0.02156616933643818
step: 40, loss: 0.000215393170947209
step: 50, loss: 0.00030645637889392674
step: 60, loss: 0.00015011502546258277
step: 70, loss: 0.00013762890012003481
step: 80, loss: 0.00013376298011280596
step: 90, loss: 5.9815207350766286e-05
step: 100, loss: 0.0011212952667847276
step: 110, loss: 0.00023841422807890922
step: 120, loss: 0.0002497639216016978
step: 130, loss: 0.0003961913753300905
step: 140, loss: 0.0023063335102051497
step: 150, loss: 0.020171958953142166
step: 160, loss: 0.00014472361363004893
step: 170, loss: 0.00019942129438277334
step: 180, loss: 0.0002558663545642048
step: 190, loss: 0.01691138744354248
step: 200, loss: 0.0004286569310352206
step: 210, loss: 9.979152673622593e-05
step: 220, loss: 0.000207486460567452
step: 230, loss: 0.000875392637681216
/usr/local/lib/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=0.9601769911504424, f1=0.9534368070953436,
best_f1=0.9534368070953436
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00012840832641813904
step: 10, loss: 6.0484991990961134e-05
step: 20, loss: 0.004971032030880451
```

```
step: 30, loss: 0.0002781510120257735
step: 40, loss: 0.0008068294846452773
step: 50, loss: 0.0001574660127516836
step: 60, loss: 0.0001665893942117691
step: 70, loss: 0.0001103335089283064
step: 80, loss: 0.00024053252127487212
step: 90, loss: 0.0002702136989682913
step: 100, loss: 0.0003977168817073107
step: 110, loss: 0.003651308361440897
step: 120, loss: 0.004561041947454214
step: 130, loss: 0.00016183915431611124
step: 140, loss: 0.00013834974379278772
step: 150, loss: 0.00015769233868923038
step: 160, loss: 5.868625157745555e-05
step: 170, loss: 0.00011352447472745553
step: 180, loss: 0.00014190253568813205
step: 190, loss: 0.00014624741743318737
step: 200, loss: 0.00010199299140367657
step: 210, loss: 0.0002676156582310796
step: 220, loss: 0.0009285097476094961
step: 230, loss: 5.563479135162197e-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=0.9573033707865168, f1=0.9513023782559457,
best_f1=0.9534368070953436
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.672803283436224e-05
step: 10, loss: 0.00016045267693698406
step: 20, loss: 0.0005548906046897173
step: 30, loss: 0.00629072729498148
step: 40, loss: 0.00010228485916741192
step: 50, loss: 0.00022146159608382732
step: 60, loss: 8.230601815739647e-05
step: 70, loss: 0.00014556283713318408
step: 80, loss: 0.0004923587548546493
step: 90, loss: 0.00010691978968679905
step: 100, loss: 0.00015930742665659636
step: 110, loss: 0.00010777737043099478
step: 120, loss: 0.0001188338064821437
step: 130, loss: 0.00022872546105645597
step: 140, loss: 0.00994477141648531
step: 150, loss: 0.00011982439900748432
step: 160, loss: 0.0009991723345592618
step: 170, loss: 5.6245102314278483e-05
step: 180, loss: 0.00011659820302156731
```

```
step: 190, loss: 0.00019628381414804608
step: 200, loss: 8.487709419569e-05
step: 210, loss: 0.0009776350343599916
step: 220, loss: 0.12423345446586609
step: 230, loss: 0.00010679682600311935
/usr/local/lib/python3.7/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=0.9600886917960089, f1=0.9533333333333333,
best_f1=0.9534368070953436
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 [00:08, 299.76it/s]
load_f1 = 0.9569060773480662
real_f1 = 0.9547960308710033
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, 358.03it/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 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: 0.7982829809188843
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4464222490787506
step: 20, loss: 0.5130362510681152
step: 30, loss: 0.46509647369384766
step: 40, loss: 0.45697081089019775
step: 50, loss: 0.4262579679489136
step: 60, loss: 0.34881386160850525
step: 70, loss: 0.24456660449504852
step: 80, loss: 0.42046743631362915
step: 90, loss: 0.3932913541793823
step: 100, loss: 0.20574498176574707
step: 110, loss: 0.13382203876972198
step: 120, loss: 0.16483166813850403
step: 130, loss: 0.051462382078170776
step: 140, loss: 0.44651874899864197
step: 150, loss: 0.03997713699936867
step: 160, loss: 0.16308239102363586
step: 170, loss: 0.3571757376194
step: 180, loss: 0.14475825428962708
step: 190, loss: 0.05860590189695358
step: 200, loss: 0.21275562047958374
step: 210, loss: 0.10961677134037018
```

```
step: 220, loss: 0.03240974247455597
step: 230, loss: 0.09209427982568741
step: 240, loss: 0.07035419344902039
step: 250, loss: 0.10267607122659683
step: 260, loss: 0.04524501785635948
step: 270, loss: 0.024730414152145386
step: 280, loss: 0.26218175888061523
step: 290, loss: 0.22570261359214783
step: 300, loss: 0.08830635249614716
step: 310, loss: 0.05208001285791397
step: 320, loss: 0.25876444578170776
step: 330, loss: 0.19156429171562195
step: 340, loss: 0.2791988253593445
step: 350, loss: 0.0571381114423275
step: 360, loss: 0.09291105717420578
step: 370, loss: 0.11560910940170288
step: 380, loss: 0.2357829511165619
step: 390, loss: 0.17992079257965088
step: 400, loss: 0.04467613250017166
step: 410, loss: 0.07972530275583267
step: 420, loss: 0.03446609154343605
step: 430, loss: 0.19771605730056763
step: 440, loss: 0.13597305119037628
step: 450, loss: 0.09677958488464355
step: 460, loss: 0.03825301304459572
step: 470, loss: 0.18752069771289825
step: 480, loss: 0.20512022078037262
step: 490, loss: 0.13242879509925842
step: 500, loss: 0.05141246318817139
step: 510, loss: 0.07536085695028305
step: 520, loss: 0.10368998348712921
step: 530, loss: 0.06921650469303131
/usr/local/lib/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=0.9031963470319634, f1=0.8968978102189782,
best_f1=0.8968978102189782
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.14758658409118652
step: 10, loss: 0.1849534511566162
step: 20, loss: 0.12472571432590485
step: 30, loss: 0.04602784290909767
```

```
step: 40, loss: 0.0070354831404984
step: 50, loss: 0.03534610569477081
step: 60, loss: 0.14841832220554352
step: 70, loss: 0.11638058722019196
step: 80, loss: 0.04458790645003319
step: 90, loss: 0.12686200439929962
step: 100, loss: 0.29792603850364685
step: 110, loss: 0.03999379277229309
step: 120, loss: 0.05961020663380623
step: 130, loss: 0.08935576677322388
step: 140, loss: 0.034836847335100174
step: 150, loss: 0.09516876935958862
step: 160, loss: 0.12636230885982513
step: 170, loss: 0.2119758576154709
step: 180, loss: 0.02471499890089035
step: 190, loss: 0.06445015966892242
step: 200, loss: 0.07098326832056046
step: 210, loss: 0.07859379053115845
step: 220, loss: 0.2956644296646118
step: 230, loss: 0.06763769686222076
step: 240, loss: 0.19837713241577148
step: 250, loss: 0.1472877711057663
step: 260, loss: 0.10624399036169052
step: 270, loss: 0.15051932632923126
step: 280, loss: 0.15498360991477966
step: 290, loss: 0.09267909824848175
step: 300, loss: 0.022046077996492386
step: 310, loss: 0.3676856458187103
step: 320, loss: 0.17332488298416138
step: 330, loss: 0.10459278523921967
step: 340, loss: 0.012971488758921623
step: 350, loss: 0.22586101293563843
step: 360, loss: 0.2055545151233673
step: 370, loss: 0.044420674443244934
step: 380, loss: 0.17351606488227844
step: 390, loss: 0.0404730886220932
step: 400, loss: 0.09949889034032822
step: 410, loss: 0.005594071000814438
step: 420, loss: 0.10594378411769867
step: 430, loss: 0.044273022562265396
step: 440, loss: 0.04025763273239136
step: 450, loss: 0.05129757523536682
step: 460, loss: 0.11632007360458374
step: 470, loss: 0.08016475290060043
step: 480, loss: 0.32457271218299866
step: 490, loss: 0.04924792796373367
step: 500, loss: 0.04077231511473656
step: 510, loss: 0.027998214587569237
```

```
step: 520, loss: 0.06820857524871826
step: 530, loss: 0.10850340873003006
/usr/local/lib/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=0.9106223678053346, f1=0.9018691588785047,
best_f1=0.9018691588785047
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03825894743204117
step: 10, loss: 0.04234534874558449
step: 20, loss: 0.1378730833530426
step: 30, loss: 0.11847293376922607
step: 40, loss: 0.00622442364692688
step: 50, loss: 0.12245883792638779
step: 60, loss: 0.02345062978565693
step: 70, loss: 0.010302943177521229
step: 80, loss: 0.16449153423309326
step: 90, loss: 0.43125614523887634
step: 100, loss: 0.01929451711475849
step: 110, loss: 0.04321666806936264
step: 120, loss: 0.016466228291392326
step: 130, loss: 0.0198192335665226
step: 140, loss: 0.02723868191242218
step: 150, loss: 0.06322763115167618
step: 160, loss: 0.005680506583303213
step: 170, loss: 0.05938377231359482
step: 180, loss: 0.055588606745004654
step: 190, loss: 0.0187629833817482
step: 200, loss: 0.012609181925654411
step: 210, loss: 0.09358619898557663
step: 220, loss: 0.012968008406460285
step: 230, loss: 0.04546431452035904
step: 240, loss: 0.14553318917751312
step: 250, loss: 0.04552246257662773
step: 260, loss: 0.01668866164982319
step: 270, loss: 0.04843959957361221
step: 280, loss: 0.046303991228342056
step: 290, loss: 0.11192066222429276
step: 300, loss: 0.033322498202323914
step: 310, loss: 0.06694579869508743
step: 320, loss: 0.23736423254013062
step: 330, loss: 0.03760208189487457
```

```
step: 340, loss: 0.008031913079321384
step: 350, loss: 0.07777071744203568
step: 360, loss: 0.029809672385454178
step: 370, loss: 0.07910262793302536
step: 380, loss: 0.0045060585252940655
step: 390, loss: 0.05904122814536095
step: 400, loss: 0.04982192441821098
step: 410, loss: 0.08997942507266998
step: 420, loss: 0.07264994829893112
step: 430, loss: 0.024079425260424614
step: 440, loss: 0.07470434904098511
step: 450, loss: 0.3387143611907959
step: 460, loss: 0.16267913579940796
step: 470, loss: 0.09984640032052994
step: 480, loss: 0.017579995095729828
step: 490, loss: 0.012687447480857372
step: 500, loss: 0.11381826549768448
step: 510, loss: 0.02992122247815132
step: 520, loss: 0.027329402044415474
step: 530, loss: 0.09111739695072174
/usr/local/lib/python3.7/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=0.9034676663542643, f1=0.8987816307403936,
best_f1=0.9018691588785047
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00949192140251398
step: 10, loss: 0.02455357275903225
step: 20, loss: 0.018866166472434998
step: 30, loss: 0.061631541699171066
step: 40, loss: 0.03715266287326813
step: 50, loss: 0.0925774872303009
step: 60, loss: 0.04223945736885071
step: 70, loss: 0.046036358922719955
step: 80, loss: 0.019735081121325493
step: 90, loss: 0.07762885093688965
step: 100, loss: 0.026903139427304268
step: 110, loss: 0.010609393939375877
step: 120, loss: 0.049230583012104034
step: 130, loss: 0.05898257717490196
step: 140, loss: 0.008153392001986504
step: 150, loss: 0.010185516439378262
step: 160, loss: 0.015793202444911003
step: 170, loss: 0.005714637227356434
step: 180, loss: 0.019362427294254303
step: 190, loss: 0.05633723363280296
```

```
step: 200, loss: 0.004069034475833178
step: 210, loss: 0.04003511741757393
step: 220, loss: 0.01116900146075378
step: 230, loss: 0.2651325762271881
step: 240, loss: 0.006894652731716633
step: 250, loss: 0.01763981394469738
step: 260, loss: 0.05506746843457222
step: 270, loss: 0.06506313383579254
step: 280, loss: 0.007697486784309149
step: 290, loss: 0.11614476889371872
step: 300, loss: 0.019400259479880333
step: 310, loss: 0.03575952351093292
step: 320, loss: 0.04865473508834839
step: 330, loss: 0.09919914603233337
step: 340, loss: 0.15374065935611725
step: 350, loss: 0.11978582292795181
step: 360, loss: 0.04501991346478462
step: 370, loss: 0.03568125516176224
step: 380, loss: 0.012294307351112366
step: 390, loss: 0.03295261040329933
step: 400, loss: 0.03676719218492508
step: 410, loss: 0.09193430095911026
step: 420, loss: 0.07059762626886368
step: 430, loss: 0.01758289337158203
step: 440, loss: 0.037335190922021866
step: 450, loss: 0.15598802268505096
step: 460, loss: 0.013091343455016613
step: 470, loss: 0.02564297243952751
step: 480, loss: 0.07587657868862152
step: 490, loss: 0.08643491566181183
step: 500, loss: 0.030564382672309875
step: 510, loss: 0.09564092010259628
step: 520, loss: 0.1388179361820221
step: 530, loss: 0.011920561082661152
/usr/local/lib/python3.7/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=0.9047399907961343, f1=0.8979779411764706,
best_f1=0.9018691588785047
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005694983061403036
step: 10, loss: 0.05972948670387268
step: 20, loss: 0.1154671311378479
step: 30, loss: 0.004450671374797821
step: 40, loss: 0.06322582066059113
step: 50, loss: 0.003744103247299790
```

```
step: 60, loss: 0.016159363090991974
step: 70, loss: 0.030527932569384575
step: 80, loss: 0.015611753799021244
step: 90, loss: 0.010125990957021713
step: 100, loss: 0.01511367503553629
step: 110, loss: 0.004830013494938612
step: 120, loss: 0.010847913101315498
step: 130, loss: 0.01847393997013569
step: 140, loss: 0.01718250662088394
step: 150, loss: 0.1071280911564827
step: 160, loss: 0.19249218702316284
step: 170, loss: 0.02645530365407467
step: 180, loss: 0.10257900506258011
step: 190, loss: 0.004902875050902367
step: 200, loss: 0.001651435042731464
step: 210, loss: 0.0028760910499840975
step: 220, loss: 0.00237189675681293
step: 230, loss: 0.0013801357708871365
step: 240, loss: 0.03192963823676109
step: 250, loss: 0.0010654230136424303
step: 260, loss: 0.044270068407058716
step: 270, loss: 0.04338373243808746
step: 280, loss: 0.13503782451152802
step: 290, loss: 0.010246182791888714
step: 300, loss: 0.09486734867095947
step: 310, loss: 0.0017548976466059685
step: 320, loss: 0.025495151057839394
step: 330, loss: 0.006351694464683533
step: 340, loss: 0.0012410851195454597
step: 350, loss: 0.005915845278650522
step: 360, loss: 0.00808770302683115
step: 370, loss: 0.014679256826639175
step: 380, loss: 0.011986649595201015
step: 390, loss: 0.012160150334239006
step: 400, loss: 0.01005606446415186
step: 410, loss: 0.0025126696564257145
step: 420, loss: 0.0009977903682738543
step: 430, loss: 0.02173933945596218
step: 440, loss: 0.008066880516707897
step: 450, loss: 0.011741871014237404
step: 460, loss: 0.0066484143026173115
step: 470, loss: 0.0016458386089652777
step: 480, loss: 0.031590305268764496
step: 490, loss: 0.011751066893339157
step: 500, loss: 0.09254276752471924
step: 510, loss: 0.08580372482538223
step: 520, loss: 0.020045340061187744
step: 530, loss: 0.019534600898623466
```

```
/usr/local/lib/python3.7/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=0.897998093422307, f1=0.8941063727839004,
best_f1=0.9018691588785047
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02261875569820404
step: 10, loss: 0.003615695284679532
step: 20, loss: 0.001492846291512251
step: 30, loss: 0.10798071324825287
step: 40, loss: 0.005292493384331465
step: 50, loss: 0.006549288984388113
step: 60, loss: 0.003564016195014119
step: 70, loss: 0.030891794711351395
step: 80, loss: 0.006126376800239086
step: 90, loss: 0.009247872978448868
step: 100, loss: 0.17160281538963318
step: 110, loss: 0.007791436742991209
step: 120, loss: 0.03728912025690079
step: 130, loss: 0.001251782989129424
step: 140, loss: 0.0035225017927587032
step: 150, loss: 0.01502310298383236
step: 160, loss: 0.0020381142385303974
step: 170, loss: 0.0045013814233243465
step: 180, loss: 0.002975250594317913
step: 190, loss: 0.04182012006640434
step: 200, loss: 0.0017221340676769614
step: 210, loss: 0.017037060111761093
step: 220, loss: 0.004888026509433985
step: 230, loss: 0.001384291215799749
step: 240, loss: 0.014991635456681252
step: 250, loss: 0.05478812754154205
step: 260, loss: 0.01636170782148838
step: 270, loss: 0.005938827060163021
step: 280, loss: 0.05668129399418831
step: 290, loss: 0.0032382067292928696
step: 300, loss: 0.013971775770187378
step: 310, loss: 0.0036689762491732836
step: 320, loss: 0.003498539561405778
step: 330, loss: 0.01680118404328823
step: 340, loss: 0.005184834823012352
step: 350, loss: 0.08557348698377609
step: 360, loss: 0.00240496126934886
step: 370, loss: 0.028351282700896263
step: 380, loss: 0.006289043463766575
step: 390, loss: 0.007012534886598587
```

```
step: 400, loss: 0.03177235648036003
step: 410, loss: 0.002664527390152216
step: 420, loss: 0.01372231263667345
step: 430, loss: 0.0009247343405149877
step: 440, loss: 0.014624989591538906
step: 450, loss: 0.012954012490808964
step: 460, loss: 0.11166241019964218
step: 470, loss: 0.19331343472003937
step: 480, loss: 0.0013070735149085522
step: 490, loss: 0.00035277154529467225
step: 500, loss: 0.003148062853142619
step: 510, loss: 0.0002703421632759273
step: 520, loss: 0.0293479785323143
step: 530, loss: 0.026543177664279938
/usr/local/lib/python3.7/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=0.900976290097629, f1=0.8984557791296208,
best_f1=0.9018691588785047
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1025884598493576
step: 10, loss: 0.021899014711380005
step: 20, loss: 0.09114135056734085
step: 30, loss: 0.010739671997725964
step: 40, loss: 0.0014450668822973967
step: 50, loss: 0.008575382642447948
step: 60, loss: 0.13863256573677063
step: 70, loss: 0.0039099715650081635
step: 80, loss: 0.0021618136670440435
step: 90, loss: 0.0023560875561088324
step: 100, loss: 0.004888542927801609
step: 110, loss: 0.0005981494323350489
step: 120, loss: 0.003330579260364175
step: 130, loss: 0.007229745388031006
step: 140, loss: 0.0012210758868604898
step: 150, loss: 0.010626249015331268
step: 160, loss: 0.03184524178504944
step: 170, loss: 0.0005580020952038467
step: 180, loss: 0.00016923858493100852
step: 190, loss: 0.00041830644477158785
step: 200, loss: 0.10008571296930313
step: 210, loss: 0.005642113275825977
step: 220, loss: 0.0012870035134255886
step: 230, loss: 0.002792528597638011
step: 240, loss: 0.007669567130506039
step: 250, loss: 0.06794651597738266
```

```
step: 260, loss: 0.0018096998101100326
step: 270, loss: 0.0008586003095842898
step: 280, loss: 0.06439966708421707
step: 290, loss: 0.012451743707060814
step: 300, loss: 0.0020789944101125
step: 310, loss: 0.0011098942486569285
step: 320, loss: 0.014999127946794033
step: 330, loss: 0.0016055674059316516
step: 340, loss: 0.0036702502984553576
step: 350, loss: 0.000509137287735939
step: 360, loss: 0.0017484967829659581
step: 370, loss: 0.052420541644096375
step: 380, loss: 0.02371697686612606
step: 390, loss: 0.0887996107339859
step: 400, loss: 0.0008282305207103491
step: 410, loss: 0.0006793743814341724
step: 420, loss: 0.07761706411838531
step: 430, loss: 0.005867110565304756
step: 440, loss: 0.0023358475882560015
step: 450, loss: 0.15369990468025208
step: 460, loss: 0.0047406842932105064
step: 470, loss: 0.07186049222946167
step: 480, loss: 0.02204899862408638
step: 490, loss: 0.0013421833282336593
step: 500, loss: 0.0023861927911639214
step: 510, loss: 0.0626869797706604
step: 520, loss: 0.011427217163145542
step: 530, loss: 0.006198153365403414
/usr/local/lib/python3.7/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=0.902867715078631, f1=0.895053166897827,
best_f1=0.9018691588785047
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0021147553343325853
step: 10, loss: 0.012118292972445488
step: 20, loss: 0.002598653081804514
step: 30, loss: 0.046417322009801865
step: 40, loss: 0.03596482798457146
step: 50, loss: 0.0071052694693320774
step: 60, loss: 0.0005788626149296761
step: 70, loss: 0.0027497904375195503
step: 80, loss: 0.0003799244877882302
step: 90, loss: 0.0014975713565945625
step: 100, loss: 0.0016718741971999407
step: 110, loss: 0.014063311740756035
```

```
step: 120, loss: 0.00292645744048059
step: 130, loss: 0.0023795857559889555
step: 140, loss: 0.00838882103562355
step: 150, loss: 0.008219695650041103
step: 160, loss: 0.004183778073638678
step: 170, loss: 0.0004660072736442089
step: 180, loss: 0.0006778939277864993
step: 190, loss: 0.02022307552397251
step: 200, loss: 0.12221097946166992
step: 210, loss: 0.07664517313241959
step: 220, loss: 0.05119811370968819
step: 230, loss: 0.014756621792912483
step: 240, loss: 0.002670933958142996
step: 250, loss: 0.0024444579612463713
step: 260, loss: 0.008775202557444572
step: 270, loss: 0.030993878841400146
step: 280, loss: 0.0031554296147078276
step: 290, loss: 0.00820474699139595
step: 300, loss: 0.0010663140565156937
step: 310, loss: 0.1537461280822754
step: 320, loss: 0.0014392659068107605
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 32768.0
step: 330, loss: 0.011940453201532364
step: 340, loss: 0.0026539878454059362
step: 350, loss: 0.0021168726962059736
step: 360, loss: 0.009972946718335152
step: 370, loss: 0.0002928411995526403
step: 380, loss: 0.013438199646770954
step: 390, loss: 0.0003107581287622452
step: 400, loss: 0.037685543298721313
step: 410, loss: 0.042224619537591934
step: 420, loss: 0.010965961031615734
step: 430, loss: 0.00027797420625574887
step: 440, loss: 0.000565855298191309
step: 450, loss: 0.00019999158394057304
step: 460, loss: 0.005684616044163704
step: 470, loss: 0.00031648701406084
step: 480, loss: 0.0012036680709570646
step: 490, loss: 0.0028630474116653204
step: 500, loss: 0.00553123839199543
step: 510, loss: 0.008456171490252018
step: 520, loss: 0.0012767510488629341
step: 530, loss: 0.0011223091278225183
/usr/local/lib/python3.7/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=0.9045368620037807, f1=0.8946360153256706,
best_f1=0.9018691588785047
```

```
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00564716849476099
step: 10, loss: 0.0056289322674274445
step: 20, loss: 0.005199304781854153
step: 30, loss: 0.00021358694357331842
step: 40, loss: 0.007544674910604954
step: 50, loss: 0.00011369844287401065
step: 60, loss: 0.0751718133687973
step: 70, loss: 0.04117821529507637
step: 80, loss: 0.0007010928820818663
step: 90, loss: 0.003998066298663616
step: 100, loss: 0.0004571712634060532
step: 110, loss: 0.0011872985633090138
step: 120, loss: 0.014522627927362919
step: 130, loss: 0.001164962537586689
step: 140, loss: 0.03517058119177818
step: 150, loss: 0.0008350281277671456
step: 160, loss: 0.0003857213305309415
step: 170, loss: 0.0016956122126430273
step: 180, loss: 0.0004570067976601422
step: 190, loss: 0.018826473504304886
step: 200, loss: 0.01827535405755043
step: 210, loss: 0.024299079552292824
step: 220, loss: 0.001977416453883052
step: 230, loss: 0.00035494109033606946
step: 240, loss: 0.0007385443896055222
step: 250, loss: 0.025908075273036957
step: 260, loss: 0.0005093482905067503
step: 270, loss: 0.0009971554391086102
step: 280, loss: 0.0007454385049641132
step: 290, loss: 0.004448758438229561
step: 300, loss: 0.006682144477963448
step: 310, loss: 0.0029678253922611475
step: 320, loss: 0.00018512620590627193
step: 330, loss: 0.0016137483762577176
step: 340, loss: 0.00033842556877061725
step: 350, loss: 0.0004195731598883867
step: 360, loss: 0.12560687959194183
step: 370, loss: 0.003002808429300785
step: 380, loss: 0.000516363128554076
step: 390, loss: 0.00045841190149076283
step: 400, loss: 0.013568051159381866
step: 410, loss: 0.0009336645598523319
step: 420, loss: 0.002737500239163637
step: 430, loss: 8.7067557615228e-05
step: 440, loss: 0.00015615567099303007
```

```
step: 450, loss: 0.00018836840172298253
step: 460, loss: 0.01116976048797369
step: 470, loss: 0.0034195471089333296
step: 480, loss: 0.00022866670042276382
step: 490, loss: 0.017325570806860924
step: 500, loss: 0.005075111053884029
step: 510, loss: 0.000851477961987257
step: 520, loss: 0.00044502259769776511
step: 530, loss: 0.011307810433208942
/usr/local/lib/python3.7/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=0.909689557855127, f1=0.8974600188146755,
best_f1=0.9018691588785047
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0014510875334963202
step: 10, loss: 5.348390550352633e-05
step: 20, loss: 0.00033563983743079007
step: 30, loss: 0.000579712213948369
step: 40, loss: 0.002069519367069006
step: 50, loss: 0.0008557431283406913
step: 60, loss: 0.00030664229416288435
step: 70, loss: 0.001862548291683197
step: 80, loss: 0.0013343878090381622
step: 90, loss: 0.00013970861618872732
step: 100, loss: 0.00041801054612733424
step: 110, loss: 0.0013092479202896357
Gradient overflow.  Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 120, loss: 0.060049451887607574
step: 130, loss: 0.00046665509914048016
step: 140, loss: 0.00018372925114817917
step: 150, loss: 0.00011535705561982468
step: 160, loss: 0.0001132856632466428
step: 170, loss: 8.702820196049288e-05
step: 180, loss: 0.0041213869117200375
step: 190, loss: 0.00425015389919281
step: 200, loss: 0.0010279413545504212
step: 210, loss: 0.0009273221949115396
step: 220, loss: 0.040195707231760025
step: 230, loss: 0.00041240412974730134
step: 240, loss: 0.00026329830870963633
step: 250, loss: 0.0007406338700093329
step: 260, loss: 0.029413679614663124
step: 270, loss: 0.018715035170316696
step: 280, loss: 0.0023499876260757446
step: 290, loss: 0.0015848601469770074
```

```
step: 300, loss: 9.948539809556678e-05
step: 310, loss: 0.0004565985582303256
step: 320, loss: 0.0008190007647499442
step: 330, loss: 0.0005142292939126492
step: 340, loss: 0.0002180187002522871
step: 350, loss: 0.0013291981304064393
step: 360, loss: 0.010360551066696644
step: 370, loss: 0.030326316133141518
step: 380, loss: 0.002050215844064951
step: 390, loss: 0.0003189879935234785
step: 400, loss: 0.001730454503558576
step: 410, loss: 0.00016337656415998936
step: 420, loss: 0.0027209350373595953
step: 430, loss: 0.00013890657282900065
step: 440, loss: 0.00028123368974775076
step: 450, loss: 0.005348957143723965
step: 460, loss: 0.006199202965945005
step: 470, loss: 0.0011373708257451653
step: 480, loss: 0.03535556420683861
step: 490, loss: 0.06211197003722191
step: 500, loss: 0.047885745763778687
step: 510, loss: 0.0009636637405492365
step: 520, loss: 0.007405717391520739
step: 530, loss: 0.00030635553412139416
/usr/local/lib/python3.7/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=0.9063809967396367, f1=0.8998127340823969,
best_f1=0.9018691588785047
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004500838927924633
step: 10, loss: 0.02629302628338337
step: 20, loss: 0.00012884955503977835
step: 30, loss: 0.002587533090263605
step: 40, loss: 0.001340128481388092
step: 50, loss: 0.0003754629287868738
step: 60, loss: 0.0007827695226296782
step: 70, loss: 0.00032670609652996063
step: 80, loss: 0.0004821764014195651
step: 90, loss: 0.023742714896798134
step: 100, loss: 0.0025149991270154715
step: 110, loss: 0.012129341252148151
step: 120, loss: 0.001681641093455255
step: 130, loss: 0.0011537373065948486
step: 140, loss: 0.00023889285512268543
step: 150, loss: 0.0008498994866386056
```

```
step: 160, loss: 0.013743972405791283
step: 170, loss: 0.10277321189641953
step: 180, loss: 0.001778391539119184
step: 190, loss: 0.0035955794155597687
step: 200, loss: 0.0014492339687421918
step: 210, loss: 0.00022756373800802976
step: 220, loss: 0.009299754165112972
step: 230, loss: 0.00369303859770298
step: 240, loss: 0.01687893457710743
step: 250, loss: 0.011287764646112919
step: 260, loss: 0.0008093130891211331
step: 270, loss: 0.0025043704081326723
step: 280, loss: 0.039190974086523056
step: 290, loss: 0.0794457346200943
step: 300, loss: 0.015268977731466293
step: 310, loss: 0.03180598095059395
step: 320, loss: 0.00975189171731472
step: 330, loss: 0.00029940297827124596
step: 340, loss: 0.0008060156833380461
step: 350, loss: 0.00048817283823154867
step: 360, loss: 0.00031645153649151325
step: 370, loss: 9.615744784241542e-05
step: 380, loss: 0.00019404084014240652
step: 390, loss: 0.010689321905374527
step: 400, loss: 5.9560901718214154e-05
step: 410, loss: 6.33394083706662e-05
step: 420, loss: 0.0016930322162806988
step: 430, loss: 0.10574033111333847
step: 440, loss: 0.0007854807772673666
step: 450, loss: 0.0023246014025062323
step: 460, loss: 0.002701259683817625
step: 470, loss: 0.0015444901073351502
step: 480, loss: 0.0018672990845516324
step: 490, loss: 0.009808644652366638
step: 500, loss: 0.00014631239173468202
step: 510, loss: 0.001237741787917912
step: 520, loss: 0.0007262483122758567
step: 530, loss: 0.000774629705119878
/usr/local/lib/python3.7/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=0.9078885214926783, f1=0.8921897460469573,
best_f1=0.9018691588785047
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.017689166590571404
step: 10, loss: 0.002758654300123453
```

```
step: 20, loss: 0.00016214372590184212
step: 30, loss: 0.0021541034802794456
step: 40, loss: 0.007923010736703873
step: 50, loss: 0.001392142497934401
step: 60, loss: 0.0013603456318378448
step: 70, loss: 0.051240477710962296
step: 80, loss: 0.0017095537623390555
step: 90, loss: 0.00031162326922640204
step: 100, loss: 0.0003878317656926811
step: 110, loss: 0.03193813934922218
step: 120, loss: 0.0048200832679867744
step: 130, loss: 0.0001300815201830119
step: 140, loss: 0.0019140883814543486
step: 150, loss: 0.00178417784627527
step: 160, loss: 0.002395977731794119
step: 170, loss: 0.0005725394585169852
step: 180, loss: 0.00011707811063388363
step: 190, loss: 4.6523498895112425e-05
step: 200, loss: 0.00017144490266218781
step: 210, loss: 0.00048115718527697027
step: 220, loss: 0.0010320550063624978
step: 230, loss: 0.0013901203637942672
step: 240, loss: 0.006509731058031321
step: 250, loss: 9.208107076119632e-05
step: 260, loss: 0.007472457364201546
step: 270, loss: 0.0021734428592026234
step: 280, loss: 0.0005063028074800968
step: 290, loss: 0.002734552603214979
step: 300, loss: 0.0009465275215916336
step: 310, loss: 0.0002366201370023191
step: 320, loss: 0.03438413888216019
step: 330, loss: 4.719717981060967e-05
step: 340, loss: 0.0026942668482661247
step: 350, loss: 0.008574465289711952
step: 360, loss: 0.06544957309961319
step: 370, loss: 0.003381578717380762
step: 380, loss: 0.0018557123839855194
step: 390, loss: 6.689751899102703e-05
step: 400, loss: 0.0006841438007541001
step: 410, loss: 0.0007656239322386682
step: 420, loss: 0.003603107761591673
step: 430, loss: 0.0009408665937371552
step: 440, loss: 0.0011350024724379182
step: 450, loss: 0.00410850066691637
step: 460, loss: 0.0005143980379216373
step: 470, loss: 0.00028481963090598583
step: 480, loss: 0.0047204080037772655
step: 490, loss: 0.0012178727192804217
```

```
step: 500, loss: 0.04734024405479431
step: 510, loss: 0.003943316638469696
step: 520, loss: 0.02699017897248268
step: 530, loss: 0.00045606750063598156
/usr/local/lib/python3.7/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=0.9052044609665428, f1=0.9004207573632538,
best_f1=0.9018691588785047
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003362427232787013
step: 10, loss: 0.0009891543304547668
step: 20, loss: 0.006844718009233475
step: 30, loss: 0.0028039757162332535
step: 40, loss: 0.0021331619936972857
step: 50, loss: 0.00027809679158963263
step: 60, loss: 0.0004339178849477321
step: 70, loss: 0.00012681067164521664
step: 80, loss: 0.003235679119825363
step: 90, loss: 0.00026608232292346656
step: 100, loss: 0.00014002232637722045
step: 110, loss: 0.00316513329744339
step: 120, loss: 0.0005847300635650754
step: 130, loss: 0.00042620583553798497
step: 140, loss: 0.00018643939984031022
step: 150, loss: 0.00016130581207107753
step: 160, loss: 0.00019262875139247626
step: 170, loss: 0.00011534024815773591
step: 180, loss: 0.00010961997759295627
step: 190, loss: 0.00036281815846450627
step: 200, loss: 0.003266807645559311
step: 210, loss: 0.010818293318152428
step: 220, loss: 0.00035568352905102074
step: 230, loss: 0.002471545711159706
step: 240, loss: 4.568430449580774e-05
step: 250, loss: 0.00783555582165718
step: 260, loss: 0.00010502606892259791
step: 270, loss: 0.0002526793396100402
step: 280, loss: 0.0005216982681304216
step: 290, loss: 0.00010251572530250996
step: 300, loss: 0.08309199661016464
step: 310, loss: 6.260585359996185e-05
step: 320, loss: 6.656361802015454e-05
step: 330, loss: 0.008517318405210972
step: 340, loss: 6.0823924286523834e-05
step: 350, loss: 0.02015695348381996
```

```
step: 360, loss: 4.7685334720881656e-05
step: 370, loss: 0.00022537459153681993
step: 380, loss: 0.0001646478776820004
step: 390, loss: 0.0027441338170319796
step: 400, loss: 8.002568210940808e-05
step: 410, loss: 0.0017210105434060097
step: 420, loss: 0.0001638496178202331
step: 430, loss: 0.00021311210002750158
step: 440, loss: 0.0007066582329571247
step: 450, loss: 0.035697903484106064
step: 460, loss: 3.017406379512977e-05
step: 470, loss: 0.00015888974303379655
step: 480, loss: 0.00015581378829665482
step: 490, loss: 0.00012976604921277612
step: 500, loss: 0.0010981657542288303
step: 510, loss: 0.0003643592062871903
step: 520, loss: 0.0031372078228741884
step: 530, loss: 4.464137236936949e-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=0.9081439393939394, f1=0.8956646021915198,
best_f1=0.9018691588785047
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.011103971861302853
step: 10, loss: 0.0002533833321649581
step: 20, loss: 0.002520930953323841
step: 30, loss: 0.0001367545046377927
step: 40, loss: 0.0016813272377476096
step: 50, loss: 0.007435955572873354
step: 60, loss: 5.764148590969853e-05
step: 70, loss: 0.00010172993643209338
step: 80, loss: 0.0008153197122737765
step: 90, loss: 3.526260843500495e-05
step: 100, loss: 0.019410571083426476
step: 110, loss: 0.0009099774179048836
step: 120, loss: 0.0005080591654404998
step: 130, loss: 0.013277019374072552
step: 140, loss: 0.001084362855181098
step: 150, loss: 0.0016927020624279976
step: 160, loss: 0.0022211463656276464
step: 170, loss: 0.029058555141091347
step: 180, loss: 0.00017117369861807674
step: 190, loss: 0.0003252696478739381
step: 200, loss: 0.0005044098361395299
step: 210, loss: 0.0011394941247999668
```

```
step: 220, loss: 0.0005045440630055964
step: 230, loss: 0.00018764031119644642
step: 240, loss: 0.029508911073207855
step: 250, loss: 0.00013873455463908613
step: 260, loss: 4.0086426452035084e-05
step: 270, loss: 0.00338375149294734
step: 280, loss: 0.0005839442601427436
step: 290, loss: 0.00014334962179418653
step: 300, loss: 0.0118851438164711
step: 310, loss: 7.801549509167671e-05
step: 320, loss: 0.0013877426972612739
step: 330, loss: 0.0004280692955944687
step: 340, loss: 8.011395402718335e-05
step: 350, loss: 0.08361838757991791
step: 360, loss: 0.004957842640578747
step: 370, loss: 0.0040078226629644632
step: 380, loss: 0.000567456241697073
step: 390, loss: 0.0003723127010744065
step: 400, loss: 8.384545799344778e-05
step: 410, loss: 0.00029917582287453115
step: 420, loss: 0.00044559402158483863
step: 430, loss: 0.03826957568526268
step: 440, loss: 8.488256571581587e-05
step: 450, loss: 0.00042476505041122437
step: 460, loss: 0.0003505288332235068
step: 470, loss: 0.0002665844513103366
step: 480, loss: 8.673420234117657e-05
step: 490, loss: 0.0002494853688403964
step: 500, loss: 0.0002464163117110729
step: 510, loss: 0.00022857241856399924
step: 520, loss: 0.00030458849505521357
step: 530, loss: 0.00014475222269953879
/usr/local/lib/python3.7/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=0.907906976744186, f1=0.9010782934833569,
best_f1=0.9018691588785047
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003322795091662556
step: 10, loss: 6.990740803303197e-05
step: 20, loss: 0.0010017832973971963
step: 30, loss: 0.05226628854870796
step: 40, loss: 5.313063593348488e-05
step: 50, loss: 0.00012790581968147308
step: 60, loss: 4.87485813209787e-05
step: 70, loss: 0.00013733140076510608
```

```
step: 80, loss: 0.00048187977517955005
step: 90, loss: 5.172291639610194e-05
step: 100, loss: 0.00012843689182773232
step: 110, loss: 0.01055108942091465
step: 120, loss: 0.0026062140241265297
step: 130, loss: 0.00624282518401742
step: 140, loss: 5.6472516007488593e-05
step: 150, loss: 0.00023452599998563528
step: 160, loss: 0.04265584796667099
step: 170, loss: 0.0023937434889376163
step: 180, loss: 0.00018533741240389645
step: 190, loss: 0.00028078255127184093
step: 200, loss: 0.00041029395652003586
step: 210, loss: 0.038416944444179535
step: 220, loss: 0.00023327778035309166
step: 230, loss: 0.0006458965362980962
step: 240, loss: 0.030704353004693985
step: 250, loss: 0.00011783543595811352
step: 260, loss: 8.200830052373931e-05
step: 270, loss: 0.0020487839356064796
step: 280, loss: 0.0004050748539157212
step: 290, loss: 6.690154259558767e-05
step: 300, loss: 0.0032447385601699352
step: 310, loss: 0.0008522473508492112
step: 320, loss: 0.00019961445959750563
step: 330, loss: 0.0023437633644789457
step: 340, loss: 0.0011092734057456255
step: 350, loss: 0.00025926329544745386
step: 360, loss: 0.0004146920400671661
step: 370, loss: 0.0001325151533819735
step: 380, loss: 0.00013743346789851785
step: 390, loss: 0.0005480347317643464
step: 400, loss: 0.0005917568923905492
step: 410, loss: 0.0003089495876338333
step: 420, loss: 0.0007494863239116967
step: 430, loss: 0.000527364551089704
step: 440, loss: 0.00036570316297002137
step: 450, loss: 0.0017207941273227334
step: 460, loss: 0.00020908673468511552
step: 470, loss: 0.0003068136575166136
step: 480, loss: 0.00033877143869176507
step: 490, loss: 0.006712194532155991
step: 500, loss: 0.0010226760059595108
step: 510, loss: 8.474996866425499e-05
step: 520, loss: 0.00023826381948310882
step: 530, loss: 0.0001066458280547522
/usr/local/lib/python3.7/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=0.9064814814814813, f1=0.8997188378631676,
best_f1=0.9018691588785047
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.")
5742it [00:14, 383.22it/s]
load_f1 = 0.909090909090909
real_f1 = 0.907906976744186
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, 381.82it/s]
```