

FMedium_50_1_2_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/Snipnext_public
      %cd Snipnext_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 | 23.86 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 117.0 MB/s
Collecting numpy==1.19.2
  Downloading numpy-1.19.2-cp37-cp37m-manylinux2010_x86_64.whl (14.5 MB)
    |                               | 14.5 MB 15.6 MB/s
Collecting regex==2019.12.20
  Downloading regex-2019.12.20-cp37-cp37m-manylinux2010_x86_64.whl (689 kB)
    |                               | 689 kB 60.9 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 53.5 MB/s
Requirement already satisfied: sklearn==0.0 in
/usr/local/lib/python3.7/dist-packages (from -r requirements.txt (line 6)) (0.0)
```

```

Collecting spacy==3.1
  Downloading
spacy-3.1.0-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (6.4 MB)
|                                     | 6.4 MB 34.1 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 | 22.72 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 4.8 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 74.8 MB/s
Collecting torch==1.4.0
  Downloading torch-1.4.0-cp37-cp37m-manylinux1_x86_64.whl (753.4 MB)
|                                     | 753.4 MB 7.0 kB/s
Collecting tqdm==4.41.0
  Downloading tqdm-4.41.0-py2.py3-none-any.whl (56 kB)
|                                     | 56 kB 5.0 MB/s
Collecting transformers==3.1.0
  Downloading transformers-3.1.0-py3-none-any.whl (884 kB)
|                                     | 884 kB 49.8 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 58.9 MB/s
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: 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: 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: murmurhash<1.1.0,>=0.28.0 in
/usr/local/lib/python3.7/dist-packages (from spacy==2.2.3->-r requirements.txt
(line 4)) (1.0.7)
Collecting catalogue<1.1.0,>=0.0.7
  Downloading catalogue-1.0.0-py2.py3-none-any.whl (7.7 kB)
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)
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 50.8 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: preshed<3.1.0,>=3.0.2 in
/usr/local/lib/python3.7/dist-packages (from spacy==2.2.3->-r requirements.txt
(line 4)) (3.0.6)
Requirement already satisfied: 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)
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)
Collecting srsly<1.1.0,>=0.1.0
  Downloading srsly-1.0.5-cp37-cp37m-manylinux2014_x86_64.whl (184 kB)
    | 184 kB 73.3 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 51.5 MB/s
Requirement already satisfied: scikit-learn in
/usr/local/lib/python3.7/dist-packages (from sklearn==0.0->-r requirements.txt
(line 6)) (1.0.2)
Requirement already satisfied: protobuf>=3.8.0 in /usr/local/lib/python3.7/dist-
packages (from tensorboardX==2.0->-r requirements.txt (line 8)) (3.17.3)
Requirement already satisfied: filelock in /usr/local/lib/python3.7/dist-

```

```

packages (from transformers==3.1.0->-r requirements.txt (line 11)) (3.7.1)
Collecting tokenizers==0.8.1.rc2
  Downloading tokenizers-0.8.1rc2-cp37-cp37m-manylinux1_x86_64.whl (3.0 MB)
    |                                     | 3.0 MB 59.3 MB/s
Collecting sacremoses
  Downloading sacremoses-0.0.53.tar.gz (880 kB)
    |                                     | 880 kB 32.3 MB/s
Requirement already satisfied: packaging in /usr/local/lib/python3.7/dist-
packages (from transformers==3.1.0->-r requirements.txt (line 11)) (21.3)
Requirement already satisfied: importlib-metadata>=0.20 in
/usr/local/lib/python3.7/dist-packages (from
catalogue<1.1.0,>=0.0.7->spacy==2.2.3->-r requirements.txt (line 4)) (4.12.0)
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-
packages (from importlib-
metadata>=0.20->catalogue<1.1.0,>=0.0.7->spacy==2.2.3->-r requirements.txt (line
4)) (3.8.1)
Requirement already satisfied: typing-extensions>=3.6.4 in
/usr/local/lib/python3.7/dist-packages (from importlib-
metadata>=0.20->catalogue<1.1.0,>=0.0.7->spacy==2.2.3->-r requirements.txt (line
4)) (4.1.1)
Requirement already satisfied: 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: 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: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-
packages (from requests<3.0.0,>=2.13.0->spacy==2.2.3->-r requirements.txt (line
4)) (2.10)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in
/usr/local/lib/python3.7/dist-packages (from packaging->transformers==3.1.0->-r
requirements.txt (line 11)) (3.0.9)
Requirement already satisfied: click in /usr/local/lib/python3.7/dist-packages
(from sacremoses->transformers==3.1.0->-r requirements.txt (line 11)) (7.1.2)
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages
(from sacremoses->transformers==3.1.0->-r requirements.txt (line 11)) (1.1.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/usr/local/lib/python3.7/dist-packages (from scikit-learn->sklearn==0.0->-r
requirements.txt (line 6)) (3.1.0)
Building wheels for collected packages: nltk, sacremoses
  Building wheel for nltk (setup.py) ... done
  Created wheel for nltk: filename=nltk-3.4.5-py3-none-any.whl size=1449920
sha256=436a49cc3f296442acb95e8d65956a12d59ce78ec5af4a58170ff889c2f7ee55
  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=d47301bdeafa950450eeaaafb7bb80d9cbd6f10ce50121cd7adddc2269ac0077
Stored in directory: /root/.cache/pip/wheels/87/39/dd/a83eeef36d0bf98e7a4d1933
a4ad2d660295a40613079bafc9
Successfully built nltk sacremoses
Installing collected packages: numpy, tqdm, srsly, regex, plac, blis,
tokenizers, thinc, sentencepiece, sacremoses, catalogue, transformers, torch,
tensorboardX, spacy, nltk, jsonlines, gensim
Attempting uninstall: numpy
  Found existing installation: numpy 1.21.6
  Uninstalling numpy-1.21.6:
    Successfully uninstalled numpy-1.21.6
Attempting uninstall: tqdm
  Found existing installation: tqdm 4.64.0
  Uninstalling tqdm-4.64.0:
    Successfully uninstalled tqdm-4.64.0
Attempting uninstall: srsly
  Found existing installation: srsly 2.4.4
  Uninstalling srsly-2.4.4:
    Successfully uninstalled srsly-2.4.4
Attempting uninstall: regex
  Found existing installation: regex 2022.6.2
  Uninstalling regex-2022.6.2:
    Successfully uninstalled regex-2022.6.2
Attempting uninstall: blis
  Found existing installation: blis 0.7.8
  Uninstalling blis-0.7.8:
    Successfully uninstalled blis-0.7.8
Attempting uninstall: thinc
  Found existing installation: thinc 8.1.0
  Uninstalling thinc-8.1.0:
    Successfully uninstalled thinc-8.1.0
Attempting uninstall: catalogue
  Found existing installation: catalogue 2.0.8
  Uninstalling catalogue-2.0.8:
    Successfully uninstalled catalogue-2.0.8
Attempting uninstall: torch
  Found existing installation: torch 1.12.0+cu113
  Uninstalling torch-1.12.0+cu113:
    Successfully uninstalled torch-1.12.0+cu113
Attempting uninstall: spacy
  Found existing installation: spacy 3.4.1
  Uninstalling spacy-3.4.1:
    Successfully uninstalled spacy-3.4.1
Attempting uninstall: nltk
  Found existing installation: nltk 3.7
```

```

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

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

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

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

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

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

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

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

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

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

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

```

```
[ ]: True
```

0.2 Install fp16 optimization

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

```
Cloning into 'apex'...
remote: Enumerating objects: 10085, done.
remote: Counting objects: 100% (201/201), done.
remote: Compressing objects: 100% (138/138), done.
remote: Total 10085 (delta 98), reused 131 (delta 59), pack-reused 9884
Receiving objects: 100% (10085/10085), 14.95 MiB | 25.95 MiB/s, done.
Resolving deltas: 100% (6902/6902), done.
/content/ditto/apex
Using pip 21.1.3 from /usr/local/lib/python3.7/dist-packages/pip (python 3.7)
Value for scheme.platlib does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.7/dist-packages
sysconfig: /usr/lib/python3.7/site-packages
Value for scheme.purelib does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.7/dist-packages
sysconfig: /usr/lib/python3.7/site-packages
Value for scheme.headers does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/include/python3.7/UNKNOWN
sysconfig: /usr/include/python3.7m/UNKNOWN
Value for scheme.scripts does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/bin
sysconfig: /usr/bin
Value for scheme.data does not match. Please report this to
<https://github.com/pypa/pip/issues/9617>
distutils: /usr/local
sysconfig: /usr
Additional context:
user = False
home = None
root = None
prefix = None
Non-user install because site-packages writable
Created temporary directory: /tmp/pip-ephem-wheel-cache-7pft01d5
Created temporary directory: /tmp/pip-req-tracker-vm3qzoqf
Initialized build tracking at /tmp/pip-req-tracker-vm3qzoqf
Created build tracker: /tmp/pip-req-tracker-vm3qzoqf
Entered build tracker: /tmp/pip-req-tracker-vm3qzoqf
```

```

Created temporary directory: /tmp/pip-install-xmwg2tye
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-yvkm56ju
  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-
vm3qzoqf'
  Running setup.py (path:/tmp/pip-req-build-yvkm56ju/setup.py) egg_info for
package from file:///content/ditto/apex
  Created temporary directory: /tmp/pip-pip-egg-info-gnvozuvv
  Running command python setup.py egg_info

  torch.__version__ = 1.4.0

  running egg_info
  creating /tmp/pip-pip-egg-info-gnvozuvv/apex.egg-info
  writing /tmp/pip-pip-egg-info-gnvozuvv/apex.egg-info/PKG-INFO
  writing dependency_links to /tmp/pip-pip-egg-info-gnvozuvv/apex.egg-
info/dependency_links.txt
  writing top-level names to /tmp/pip-pip-egg-info-gnvozuvv/apex.egg-
info/top_level.txt
  writing manifest file '/tmp/pip-pip-egg-info-gnvozuvv/apex.egg-
info/SOURCES.txt'
  adding license file 'LICENSE'
  writing manifest file '/tmp/pip-pip-egg-info-gnvozuvv/apex.egg-
info/SOURCES.txt'
  Source in /tmp/pip-req-build-yvkm56ju 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-vm3qzoqf'
Created temporary directory: /tmp/pip-unpack-oy3ovztu
Building wheels for collected packages: apex
  Created temporary directory: /tmp/pip-wheel-aw0ct4z9
  Building wheel for apex (setup.py) ... Destination directory:
/tmp/pip-wheel-aw0ct4z9
  Running command /usr/bin/python3 -u -c 'import io, os, sys, setuptools,
tokenize; sys.argv[0] = ''"/tmp/pip-req-build-yvkm56ju/setup.py''';

```



```
__file__='''/tmp/pip-req-build-yvkm56ju/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-aw0ct4z9
```

```
torch.__version__ = 1.4.0
```

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

```

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

```

```

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

```

```

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

```

```

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

```

```

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

```

```

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

```

```

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

```



```

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

```

```

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

```

```

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

```

```

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

```

```

build/bdist.linux-x86_64/wheel/apex/fp16_utils
  copying build/lib/apex/fp16_utils/fp16_optimizer.py ->
build/bdist.linux-x86_64/wheel/apex/fp16_utils
  creating build/bdist.linux-x86_64/wheel/apex/fused_dense
  copying build/lib/apex/fused_dense/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/fused_dense
  copying build/lib/apex/fused_dense/fused_dense.py ->
build/bdist.linux-x86_64/wheel/apex/fused_dense
  creating build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/optimized_sync_batchnorm.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/optimized_sync_batchnorm_kernel.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/sync_batchnorm_kernel.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/multiproc.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/distributed.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/LARC.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  copying build/lib/apex/parallel/sync_batchnorm.py ->
build/bdist.linux-x86_64/wheel/apex/parallel
  creating build/bdist.linux-x86_64/wheel/apex/mlp
  copying build/lib/apex/mlp/__init__.py ->
build/bdist.linux-x86_64/wheel/apex/mlp
  copying build/lib/apex/mlp/mlp.py -> build/bdist.linux-x86_64/wheel/apex/mlp
running install_egg_info
running egg_info
creating apex.egg-info
writing apex.egg-info/PKG-INFO
writing dependency_links to apex.egg-info/dependency_links.txt
writing top-level names to apex.egg-info/top_level.txt
writing manifest file 'apex.egg-info/SOURCES.txt'
adding license file 'LICENSE'
writing manifest file 'apex.egg-info/SOURCES.txt'
Copying apex.egg-info to build/bdist.linux-x86_64/wheel/apex-0.1-py3.7.egg-
info
running install_scripts
adding license file "LICENSE" (matched pattern "LICEN[CS]E*")
creating build/bdist.linux-x86_64/wheel/apex-0.1.dist-info/WHEEL
creating '/tmp/pip-wheel-aw0ct4z9/apex-0.1-py3-none-any.whl' and adding
'build/bdist.linux-x86_64/wheel' to it
  adding 'apex/__init__.py'
  adding 'apex/_autocast_utils.py'
  adding 'apex/RNN/RNNBackend.py'

```

```

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

```

```

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

```

```

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

```



```

adding 'apex/transformer/testing/standalone_transformer_lm.py'
adding 'apex-0.1.dist-info/LICENSE'
adding 'apex-0.1.dist-info/METADATA'
adding 'apex-0.1.dist-info/WHEEL'
adding 'apex-0.1.dist-info/top_level.txt'
adding 'apex-0.1.dist-info/RECORD'
removing build/bdist.linux-x86_64/wheel
done
Created wheel for apex: filename=apex-0.1-py3-none-any.whl size=298114
sha256=ba9c2b8a591a0053a9f379082c2e64590b79e34c986a1e43b01bf27fab6a13fe
Stored in directory: /tmp/pip-ephem-wheel-cache-7pft01d5/wheels/6b/22/a2/16655
26ee3c3061243260fd522525eca31398e04bfa5ad7e1f
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-vm3qzoqf'
/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 5.0 MB/s
Collecting awscli
  Downloading awscli-1.25.46-py3-none-any.whl (3.9 MB)
    |                                     | 3.9 MB 49.6 MB/s
Collecting botocore==1.27.46
  Downloading botocore-1.27.46-py3-none-any.whl (9.0 MB)
    |                                     | 9.0 MB 36.0 MB/s
Requirement already satisfied: PyYAML<5.5,>=3.10 in
/usr/local/lib/python3.7/dist-packages (from awscli) (3.13)
Collecting rsa<4.8,>=3.1.2
  Downloading rsa-4.7.2-py3-none-any.whl (34 kB)
Collecting colorama<0.4.5,>=0.2.5
  Downloading colorama-0.4.4-py2.py3-none-any.whl (16 kB)
Collecting s3transfer<0.7.0,>=0.6.0
  Downloading s3transfer-0.6.0-py3-none-any.whl (79 kB)
    |                                     | 79 kB 9.8 MB/s
Collecting docutils<0.17,>=0.10
  Downloading docutils-0.16-py2.py3-none-any.whl (548 kB)
    |                                     | 548 kB 60.4 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 18 kB/s
Requirement already satisfied: typing-extensions in
/usr/local/lib/python3.7/dist-packages (from torch==1.7.1) (4.1.1)
Requirement already satisfied: numpy in /usr/local/lib/python3.7/dist-packages
(from torch==1.7.1) (1.19.2)
Installing collected packages: torch
  Attempting uninstall: torch
    Found existing installation: torch 1.4.0

```

```

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

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

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

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

```

0.3 importa config.json

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

```

Cloning into 'IC'...
remote: Enumerating objects: 982, done.
remote: Counting objects: 100% (52/52), done.
remote: Compressing objects: 100% (48/48), done.
remote: Total 982 (delta 16), reused 7 (delta 0), pack-reused 930
Receiving objects: 100% (982/982), 251.89 MiB | 31.00 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/FMedium_50_1_2/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, 359kB/s]

Downloading: 100% 232k/232k [00:00<00:00, 349kB/s]

Downloading: 100% 268M/268M [00:04<00:00, 63.4MB/s]

step: 0, loss: 0.8605033159255981

epoch 1: dev_f1=0.2666666666666667, f1=0.2666666666666667,

best_f1=0.2666666666666667

step: 0, loss: 0.37809842824935913

epoch 2: dev_f1=0.3010752688172043, f1=0.2916666666666667,

best_f1=0.2916666666666667

step: 0, loss: 0.3483726978302002

epoch 3: dev_f1=0.2666666666666667, f1=0.2666666666666667,

best_f1=0.2916666666666667

```

step: 0, loss: 0.3688414990901947
epoch 4: dev_f1=0.5306122448979592, f1=0.3870967741935483,
best_f1=0.3870967741935483
step: 0, loss: 0.29190611839294434
epoch 5: dev_f1=0.5490196078431372, f1=0.4067796610169491,
best_f1=0.4067796610169491
step: 0, loss: 0.3167645037174225
epoch 6: dev_f1=0.5405405405405405, f1=0.42553191489361697,
best_f1=0.4067796610169491
step: 0, loss: 0.2359190732240677
epoch 7: dev_f1=0.5217391304347825, f1=0.3773584905660377,
best_f1=0.4067796610169491
step: 0, loss: 0.3528596758842468
epoch 8: dev_f1=0.5217391304347825, f1=0.43999999999999995,
best_f1=0.4067796610169491
step: 0, loss: 0.18669261038303375
epoch 9: dev_f1=0.52, f1=0.46153846153846156, best_f1=0.4067796610169491
step: 0, loss: 0.22426925599575043
epoch 10: dev_f1=0.5833333333333334, f1=0.36363636363636365,
best_f1=0.36363636363636365
step: 0, loss: 0.2659190893173218
epoch 11: dev_f1=0.6153846153846153, f1=0.375, best_f1=0.375
step: 0, loss: 0.20489777624607086
epoch 12: dev_f1=0.6451612903225806, f1=0.4864864864864865,
best_f1=0.4864864864864865
step: 0, loss: 0.16299884021282196
epoch 13: dev_f1=0.689655172413793, f1=0.4615384615384615,
best_f1=0.4615384615384615
step: 0, loss: 0.17777344584465027
epoch 14: dev_f1=0.6451612903225806, f1=0.45, best_f1=0.4615384615384615
step: 0, loss: 0.24326741695404053
epoch 15: dev_f1=0.6451612903225806, f1=0.45, best_f1=0.4615384615384615
mv: cannot stat '*_dev.pt': No such file or directory
Traceback (most recent call last):
  File "matcher.py", line 315, in <module>
    hp.lm, hp.use_gpu, hp.fp16)
  File "matcher.py", line 274, in load_model
    raise ModelNotFoundError(checkpoint)
ditto_light.exceptions.ModelNotFoundError: Model
checkpoints/Structured/Beer/model.pt was not found

```

1.1.2 DBLP-ACM - Running the matcher

```

[ ]: # Step 1:
      # train the wdc model with the save_model flag on
      # it will produce two files *_dev.pt and *_test.pt which are the

```

```

# best checkpoints on the dev set and the test set.
!CUDA_VISIBLE_DEVICES=0 python train_ditto.py \
  --task Structured/DBLP-ACM \
  --batch_size 32 \
  --max_len 256 \
  --lr 3e-5 \
  --n_epochs 15 \
  --finetuning \
  --lm distilbert \
  --fp16 \
  --save_model \
  --da entry_swap \
  --da span_shuffle \
  --summarize

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

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

```

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

Defaults for this optimization level are:

```

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

```

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

After processing overrides, optimization options are:

```

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

```

```

loss_scale          : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8177670240402222
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.47653302550315857
step: 20, loss: 0.5803341269493103
step: 30, loss: 0.43055468797683716
step: 40, loss: 0.1478041261434555
step: 50, loss: 0.19769889116287231
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 60, loss: 0.4840944707393646
step: 70, loss: 0.1930822730064392
step: 80, loss: 0.1743575632572174
step: 90, loss: 0.018513429909944534
step: 100, loss: 0.08200615644454956
step: 110, loss: 0.017205823212862015
step: 120, loss: 0.08063864707946777
step: 130, loss: 0.01051569264382124
step: 140, loss: 0.02078942023217678
step: 150, loss: 0.08464781939983368
step: 160, loss: 0.1983843892812729
step: 170, loss: 0.005922598298639059
step: 180, loss: 0.02682497724890709
step: 190, loss: 0.004964197054505348
step: 200, loss: 0.002528608776628971
step: 210, loss: 0.007808880880475044
step: 220, loss: 0.010411891154944897
step: 230, loss: 0.06437914073467255
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9733333333333333, f1=0.9685393258426966,
best_f1=0.9685393258426966
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012818925082683563
step: 10, loss: 0.0011346081737428904
step: 20, loss: 0.0014670944074168801
step: 30, loss: 0.0057631297968328
step: 40, loss: 0.007193882949650288
step: 50, loss: 0.001969036180526018
step: 60, loss: 0.0063567981123924255
step: 70, loss: 0.12323005497455597
step: 80, loss: 0.0555175356566906
step: 90, loss: 0.15642809867858887
step: 100, loss: 0.008862190879881382
step: 110, loss: 0.042332038283348083
step: 120, loss: 0.08576291054487228
step: 130, loss: 0.02255924977362156
step: 140, loss: 0.11372863501310349
step: 150, loss: 0.009250590577721596
step: 160, loss: 0.035656195133924484
step: 170, loss: 0.10668187588453293
step: 180, loss: 0.0048856050707399845
step: 190, loss: 0.06252799183130264
step: 200, loss: 0.0033455570228397846
step: 210, loss: 0.04362136870622635
step: 220, loss: 0.0010238747345283628
step: 230, loss: 0.01063282135874033
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9751693002257337, f1=0.9703872437357631,
best_f1=0.9703872437357631
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02539941854774952
step: 10, loss: 0.04675262048840523
step: 20, loss: 0.0028078227769583464
step: 30, loss: 0.01139899156987667
step: 40, loss: 0.008711760863661766

```

```

step: 50, loss: 0.005659281276166439
step: 60, loss: 0.038636334240436554
step: 70, loss: 0.0012454816605895758
step: 80, loss: 0.002508278237655759
step: 90, loss: 0.005059162620455027
step: 100, loss: 0.0034024915657937527
step: 110, loss: 0.004047020338475704
step: 120, loss: 0.0013788845390081406
step: 130, loss: 0.0027706571854650974
step: 140, loss: 0.03227119520306587
step: 150, loss: 0.012469133362174034
step: 160, loss: 0.002444422570988536
step: 170, loss: 0.010212581604719162
step: 180, loss: 0.0019757382106035948
step: 190, loss: 0.0032516091596335173
step: 200, loss: 0.0009851537179201841
step: 210, loss: 0.02394862473011017
step: 220, loss: 0.0012307842262089252
step: 230, loss: 0.02175178751349449
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9932584269662922, f1=0.9786276715410572,
best_f1=0.9786276715410572
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0018654776504263282
step: 10, loss: 0.0007586188148707151
step: 20, loss: 0.002150446642190218
step: 30, loss: 0.00018879779963754117
step: 40, loss: 0.052344392985105515
step: 50, loss: 0.006267173681408167
step: 60, loss: 0.0009061471209861338
step: 70, loss: 0.006181669887155294
step: 80, loss: 0.0829647108912468
step: 90, loss: 0.15091951191425323
step: 100, loss: 0.01153968833386898
step: 110, loss: 0.004665350075811148
step: 120, loss: 0.005437116138637066
step: 130, loss: 0.008490272797644138
step: 140, loss: 0.010746756568551064
step: 150, loss: 0.0008247552905231714
step: 160, loss: 0.0018751174211502075

```

```

step: 170, loss: 0.007762019522488117
step: 180, loss: 0.03047678992152214
step: 190, loss: 0.006620234344154596
step: 200, loss: 0.0029985050205141306
step: 210, loss: 0.000567488488741219
step: 220, loss: 0.0004166890576016158
step: 230, loss: 0.00026040684315375984
/usr/local/lib/python3.7/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=0.9853768278965129, f1=0.9832026875699889,
best_f1=0.9786276715410572
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00043935765279456973
step: 10, loss: 0.003510465379804373
step: 20, loss: 0.000444949371740222
step: 30, loss: 0.0001699363929219544
step: 40, loss: 0.0009564850479364395
step: 50, loss: 0.000631253351457417
step: 60, loss: 0.00043816131073981524
step: 70, loss: 0.000424716534325853
step: 80, loss: 0.0008984270971268415
step: 90, loss: 0.003705447306856513
step: 100, loss: 0.005321203265339136
step: 110, loss: 0.0014661942841485143
step: 120, loss: 0.05500144883990288
step: 130, loss: 0.058532025665044785
step: 140, loss: 0.0003922645701095462
step: 150, loss: 0.00109310622792691
step: 160, loss: 0.0029481034725904465
step: 170, loss: 0.14408427476882935
step: 180, loss: 0.0017166383331641555
step: 190, loss: 0.0006057019345462322
step: 200, loss: 0.012771505862474442
step: 210, loss: 0.00015673929010517895
step: 220, loss: 0.00036760696093551815
step: 230, loss: 0.000360829260898754
/usr/local/lib/python3.7/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=0.9910112359550561, f1=0.9832026875699889,
best_f1=0.9786276715410572
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006174965528771281

```

```

step: 10, loss: 0.0009121844777837396
step: 20, loss: 0.00024424269213341177
step: 30, loss: 0.00032329626264981925
step: 40, loss: 0.0017939693061634898
step: 50, loss: 0.001029717386700213
step: 60, loss: 0.03357701376080513
step: 70, loss: 0.0003202004008926451
step: 80, loss: 0.000596254481934011
step: 90, loss: 0.0034050652757287025
step: 100, loss: 0.00020833080634474754
step: 110, loss: 0.06578987836837769
step: 120, loss: 0.00019386695930734277
step: 130, loss: 0.010670975781977177
step: 140, loss: 0.03945178538560867
step: 150, loss: 0.0002455174981150776
step: 160, loss: 0.014780988916754723
step: 170, loss: 0.0014943117275834084
step: 180, loss: 0.0019475348526611924
step: 190, loss: 0.004829561337828636
step: 200, loss: 0.0012647656258195639
step: 210, loss: 0.0005800917278975248
step: 220, loss: 0.0007001861813478172
step: 230, loss: 0.00017080128600355238
/usr/local/lib/python3.7/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=0.9886621315192743, f1=0.9830890642615557,
best_f1=0.9786276715410572
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06851796805858612
step: 10, loss: 0.00010541062511038035
step: 20, loss: 0.00015837835962884128
step: 30, loss: 0.00017218809807673097
step: 40, loss: 0.02468627318739891
step: 50, loss: 0.0002932837523985654
step: 60, loss: 0.002499439287930727
step: 70, loss: 0.000620709266513586
step: 80, loss: 8.850958693074062e-05
step: 90, loss: 0.005064358003437519
step: 100, loss: 0.002479062881320715
step: 110, loss: 0.006513459607958794
step: 120, loss: 0.00028540624771267176
step: 130, loss: 0.0019102290971204638
step: 140, loss: 0.00010456879681441933
step: 150, loss: 0.0013886565575376153
step: 160, loss: 0.00027475968818180263

```

```

step: 170, loss: 0.00011983658623648807
step: 180, loss: 0.00033080161665566266
step: 190, loss: 0.00046099224709905684
step: 200, loss: 0.0006554732681252062
step: 210, loss: 6.508463411591947e-05
step: 220, loss: 0.0007072255830280483
step: 230, loss: 0.027097495272755623
/usr/local/lib/python3.7/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=0.9864559819413092, f1=0.9820627802690582,
best_f1=0.9786276715410572
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.025828110054135323
step: 10, loss: 0.033622339367866516
step: 20, loss: 0.0001415909791830927
step: 30, loss: 0.00017127681348938495
step: 40, loss: 0.00011746522795874625
step: 50, loss: 0.0011051947949454188
step: 60, loss: 0.0001585406280355528
step: 70, loss: 0.00012424714805092663
step: 80, loss: 0.00012745390995405614
step: 90, loss: 0.0006453428650274873
step: 100, loss: 0.0001733323442749679
step: 110, loss: 0.00042826292337849736
step: 120, loss: 0.0003524089406710118
step: 130, loss: 0.023440197110176086
step: 140, loss: 0.0014633401297032833
step: 150, loss: 0.001871301094070077
step: 160, loss: 0.00020724418573081493
step: 170, loss: 0.000782150193117559
step: 180, loss: 0.005358040798455477
step: 190, loss: 0.004022815264761448
step: 200, loss: 0.0016703654546290636
step: 210, loss: 0.0001456262543797493
step: 220, loss: 0.00018654949963092804
step: 230, loss: 0.020915264263749123
/usr/local/lib/python3.7/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=0.9863325740318906, f1=0.9749430523917996,
best_f1=0.9786276715410572
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001485277432948351

```

```

step: 10, loss: 0.0005139340064488351
step: 20, loss: 0.004523284267634153
step: 30, loss: 0.0002209645026596263
step: 40, loss: 0.00036944958264939487
step: 50, loss: 0.0002386247942922637
step: 60, loss: 0.0005331492284312844
step: 70, loss: 0.0004810365498997271
step: 80, loss: 0.0647817850112915
step: 90, loss: 0.002145488979294896
step: 100, loss: 0.00042221040348522365
step: 110, loss: 0.00025707861641421914
step: 120, loss: 0.042954131960868835
step: 130, loss: 0.0001320161100011319
step: 140, loss: 0.0005841002566739917
step: 150, loss: 7.956029003253207e-05
step: 160, loss: 0.06267859786748886
step: 170, loss: 0.0016498514451086521
step: 180, loss: 0.0006663151434622705
step: 190, loss: 0.001550276530906558
step: 200, loss: 0.00023338774917647243
step: 210, loss: 0.00027955579571425915
step: 220, loss: 0.057010117918252945
step: 230, loss: 0.016056381165981293
/usr/local/lib/python3.7/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=0.984090909090909, f1=0.9727272727272728,
best_f1=0.9786276715410572
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0012411525240167975
step: 10, loss: 0.00014469993766397238
step: 20, loss: 0.00014511565677821636
step: 30, loss: 0.0002229693200206384
step: 40, loss: 6.598087929887697e-05
step: 50, loss: 0.0013417932204902172
step: 60, loss: 0.04475465044379234
step: 70, loss: 0.00021027651382610202
step: 80, loss: 0.0025444990023970604
step: 90, loss: 0.01513691246509552
step: 100, loss: 0.0002544767630752176
step: 110, loss: 0.00015660551434848458
step: 120, loss: 0.019901404157280922
step: 130, loss: 0.0001452750148018822
step: 140, loss: 0.00366220367141068
step: 150, loss: 0.001084527699276805
step: 160, loss: 0.0016071537975221872

```

```

step: 170, loss: 0.00011427349818404764
step: 180, loss: 0.00010879190085688606
step: 190, loss: 0.00010826412471942604
step: 200, loss: 0.0004387189110275358
step: 210, loss: 0.00010025005758507177
step: 220, loss: 0.0012339549139142036
step: 230, loss: 0.00011925049329875037
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 10: dev_f1=0.9886621315192743, f1=0.9773755656108598,
best_f1=0.9786276715410572
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.077019836287946e-05
step: 10, loss: 0.00016288086771965027
step: 20, loss: 8.143940794980153e-05
step: 30, loss: 0.000203630916075781
step: 40, loss: 0.05211849510669708
step: 50, loss: 0.0006381623679772019
step: 60, loss: 0.00017858938372228295
step: 70, loss: 0.000300033570965752
step: 80, loss: 0.0001885731180664152
step: 90, loss: 0.0004481924115680158
step: 100, loss: 0.002068696543574333
step: 110, loss: 0.00020697194850072265
step: 120, loss: 0.0006738989031873643
step: 130, loss: 6.962852057768032e-05
step: 140, loss: 8.986660395748913e-05
step: 150, loss: 4.467400867724791e-05
step: 160, loss: 0.0002638653095345944
step: 170, loss: 0.0013473337749019265
step: 180, loss: 0.0008078778628259897
step: 190, loss: 0.0018039786955341697
step: 200, loss: 0.00011476716463221237
step: 210, loss: 8.710876863915473e-05
step: 220, loss: 0.00013140923692844808
step: 230, loss: 0.016130957752466202
/usr/local/lib/python3.7/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=0.9886621315192743, f1=0.9796380090497738,
best_f1=0.9786276715410572
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007082093041390181

```

```

step: 10, loss: 0.003635734086856246
step: 20, loss: 0.00011297743913019076
step: 30, loss: 0.0001466000103391707
step: 40, loss: 5.581498407991603e-05
step: 50, loss: 7.928629929665476e-05
step: 60, loss: 0.00824146717786789
step: 70, loss: 0.0004126105341129005
step: 80, loss: 5.5114196584327146e-05
step: 90, loss: 0.00012778356904163957
step: 100, loss: 0.00561614241451025
step: 110, loss: 6.0884092818014324e-05
step: 120, loss: 0.00010420790931675583
step: 130, loss: 0.0001965617120731622
step: 140, loss: 0.0031661363318562508
step: 150, loss: 6.444816244766116e-05
step: 160, loss: 0.017694832757115364
step: 170, loss: 0.00014234492846298963
step: 180, loss: 0.0010885205119848251
step: 190, loss: 0.00039061385905370116
step: 200, loss: 0.01720276288688183
step: 210, loss: 5.769883136963472e-05
step: 220, loss: 0.036298803985118866
step: 230, loss: 6.911204400239512e-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=0.9886621315192743, f1=0.9807909604519773,
best_f1=0.9786276715410572
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04079844802618027
step: 10, loss: 0.0002631599782034755
step: 20, loss: 0.00012607508688233793
step: 30, loss: 0.019872792065143585
step: 40, loss: 7.544364052591845e-05
step: 50, loss: 6.370728806359693e-05
step: 60, loss: 0.00018648739205673337
step: 70, loss: 5.157242776476778e-05
step: 80, loss: 0.00019865580543410033
step: 90, loss: 3.785895387409255e-05
step: 100, loss: 0.00010873647988773882
step: 110, loss: 7.932332664495334e-05
step: 120, loss: 8.184475882444531e-05
step: 130, loss: 0.00012570095714181662
step: 140, loss: 6.506409408757463e-05
step: 150, loss: 0.0258315559476614
step: 160, loss: 3.966362419305369e-05

```



```

step: 170, loss: 7.709822966717184e-05
step: 180, loss: 0.0001233516086358577
step: 190, loss: 2.695489820325747e-05
step: 200, loss: 4.800395981874317e-05
step: 210, loss: 0.0016762467566877604
step: 220, loss: 4.4155494833830744e-05
step: 230, loss: 2.7685069653671235e-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=0.9864253393665158, f1=0.9819413092550789,
best_f1=0.9786276715410572
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00033895979868248105
step: 10, loss: 6.289055454544723e-05
step: 20, loss: 0.018115947023034096
step: 30, loss: 0.0003375436645001173
step: 40, loss: 6.027509880368598e-05
step: 50, loss: 0.00016339322610292584
step: 60, loss: 0.06903138011693954
step: 70, loss: 5.040454925619997e-05
step: 80, loss: 0.0006803221767768264
step: 90, loss: 6.058693179511465e-05
step: 100, loss: 0.006218113005161285
step: 110, loss: 9.845043678069487e-05
step: 120, loss: 0.0001402088237227872
step: 130, loss: 0.0001291390071855858
step: 140, loss: 4.962451566825621e-05
step: 150, loss: 6.493656837847084e-05
step: 160, loss: 3.0817223887424916e-05
step: 170, loss: 5.3220184781821445e-05
step: 180, loss: 6.729440065100789e-05
step: 190, loss: 0.005038454197347164
step: 200, loss: 4.2490719351917505e-05
step: 210, loss: 0.002798468107357621
step: 220, loss: 0.00038739386945962906
step: 230, loss: 2.2246409571380354e-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=0.9852440408626559, f1=0.9774266365688488,
best_f1=0.9786276715410572
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.886568785877898e-05

```

```

step: 10, loss: 0.0001924248063005507
step: 20, loss: 0.0008529350161552429
step: 30, loss: 7.057376205921173e-05
step: 40, loss: 0.00039194533019326627
step: 50, loss: 6.258672510739416e-05
step: 60, loss: 8.190179505618289e-05
step: 70, loss: 5.918144233874045e-05
step: 80, loss: 0.017671503126621246
step: 90, loss: 3.326274600112811e-05
step: 100, loss: 9.51798283495009e-05
step: 110, loss: 4.841355621465482e-05
step: 120, loss: 0.00011419641668908298
step: 130, loss: 0.00011971587082371116
step: 140, loss: 7.378813461400568e-05
step: 150, loss: 0.0005543319857679307
step: 160, loss: 0.163773775100708
step: 170, loss: 3.108909368165769e-05
step: 180, loss: 5.1594375690910965e-05
step: 190, loss: 0.01017586700618267
step: 200, loss: 5.8779936807695776e-05
step: 210, loss: 0.0319247841835022
step: 220, loss: 5.9168818552279845e-05
step: 230, loss: 0.00010059621854452416
/usr/local/lib/python3.7/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=0.9852440408626559, f1=0.9774266365688488,
best_f1=0.9786276715410572
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

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

```

```

loss_scale          : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2473it [00:10, 235.35it/s]
load_f1 = 0.9921259842519685
real_f1 = 0.9910313901345291
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, 254.26it/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, 389kB/s]  
Downloading: 100% 232k/232k [00:00<00:00, 3.81MB/s]  
Downloading: 100% 268M/268M [00:09<00:00, 27.6MB/s]  
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32  
master weights.
```

Defaults for this optimization level are:

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

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

After processing overrides, optimization options are:

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

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

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

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

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

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

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

```
step: 0, loss: 0.7942638993263245
```

```
step: 10, loss: 0.41494306921958923
```

```
step: 20, loss: 0.4708937704563141
```

step: 30, loss: 0.38706547021865845
step: 40, loss: 0.2975634038448334
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 50, loss: 0.25888314843177795
step: 60, loss: 0.2655530273914337
step: 70, loss: 0.10779835283756256
step: 80, loss: 0.1403403878211975
step: 90, loss: 0.32250356674194336
step: 100, loss: 0.2787337303161621
step: 110, loss: 0.09484002739191055
step: 120, loss: 0.10526324808597565
step: 130, loss: 0.0313292071223259
step: 140, loss: 0.1745583415031433
step: 150, loss: 0.023143857717514038
step: 160, loss: 0.17430059611797333
step: 170, loss: 0.11957123875617981
step: 180, loss: 0.18513409793376923
step: 190, loss: 0.04387404024600983
step: 200, loss: 0.09635055065155029
step: 210, loss: 0.14454708993434906
step: 220, loss: 0.0781988874077797
step: 230, loss: 0.08446910977363586
step: 240, loss: 0.04147029668092728
step: 250, loss: 0.08178019523620605
step: 260, loss: 0.017659801989793777
step: 270, loss: 0.023093944415450096
step: 280, loss: 0.0889972671866417
step: 290, loss: 0.05499959737062454
step: 300, loss: 0.06583527475595474
step: 310, loss: 0.12628273665905
step: 320, loss: 0.0693998634815216
step: 330, loss: 0.08628088980913162
step: 340, loss: 0.17010283470153809
step: 350, loss: 0.06298727542161942
step: 360, loss: 0.0762130543589592
step: 370, loss: 0.15603715181350708
step: 380, loss: 0.15873979032039642
step: 390, loss: 0.02448088303208351
step: 400, loss: 0.03344448283314705
step: 410, loss: 0.04031778499484062
step: 420, loss: 0.09912081807851791
step: 430, loss: 0.14701522886753082
step: 440, loss: 0.11553918570280075
step: 450, loss: 0.023132657632231712
step: 460, loss: 0.04449824243783951
step: 470, loss: 0.1970556676387787
step: 480, loss: 0.32216230034828186
step: 490, loss: 0.05721587315201759

```

step: 500, loss: 0.009280765429139137
step: 510, loss: 0.06228819489479065
step: 520, loss: 0.09248702973127365
step: 530, loss: 0.16941839456558228
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9224376731301939, f1=0.9247706422018349,
best_f1=0.9247706422018349
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02264818176627159
step: 10, loss: 0.11308536678552628
step: 20, loss: 0.11084746569395065
step: 30, loss: 0.08390074968338013
step: 40, loss: 0.01939178816974163
step: 50, loss: 0.24069799482822418
step: 60, loss: 0.19343078136444092
step: 70, loss: 0.13926680386066437
step: 80, loss: 0.02532639540731907
step: 90, loss: 0.023495744913816452
step: 100, loss: 0.1941896378993988
step: 110, loss: 0.02734253741800785
step: 120, loss: 0.03655384108424187
step: 130, loss: 0.01971685141324997
step: 140, loss: 0.019694507122039795
step: 150, loss: 0.04340636357665062
step: 160, loss: 0.035401850938797
step: 170, loss: 0.05671051889657974
step: 180, loss: 0.05011541023850441
step: 190, loss: 0.010545216500759125
step: 200, loss: 0.017901333048939705
step: 210, loss: 0.013501692563295364
step: 220, loss: 0.12976382672786713
step: 230, loss: 0.011180216446518898
step: 240, loss: 0.09043984860181808
step: 250, loss: 0.0629386156797409
step: 260, loss: 0.014720543287694454
step: 270, loss: 0.16536761820316315
step: 280, loss: 0.14250041544437408
step: 290, loss: 0.12318934500217438
step: 300, loss: 0.05041610449552536
step: 310, loss: 0.13267047703266144

```

```

step: 320, loss: 0.10712742805480957
step: 330, loss: 0.10518710315227509
step: 340, loss: 0.07075381278991699
step: 350, loss: 0.06125089153647423
step: 360, loss: 0.1156829446554184
step: 370, loss: 0.0015278542414307594
step: 380, loss: 0.09971156716346741
step: 390, loss: 0.03473670035600662
step: 400, loss: 0.10060688853263855
step: 410, loss: 0.0029425269458442926
step: 420, loss: 0.02539648488163948
step: 430, loss: 0.01581326313316822
step: 440, loss: 0.048702117055654526
step: 450, loss: 0.038124021142721176
step: 460, loss: 0.3410027325153351
step: 470, loss: 0.06008133292198181
step: 480, loss: 0.25540345907211304
step: 490, loss: 0.04317926615476608
step: 500, loss: 0.024556173011660576
step: 510, loss: 0.06596110761165619
step: 520, loss: 0.04606441408395767
step: 530, loss: 0.16063928604125977
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.925497454881999, f1=0.9242144177449169,
best_f1=0.9242144177449169
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0022013233974575996
step: 10, loss: 0.194450244307518
step: 20, loss: 0.2676335871219635
step: 30, loss: 0.1870640069246292
step: 40, loss: 0.00561843765899539
step: 50, loss: 0.004300371278077364
step: 60, loss: 0.0038839641492813826
step: 70, loss: 0.07624446600675583
step: 80, loss: 0.03575309365987778
step: 90, loss: 0.10377368330955505
step: 100, loss: 0.02385229803621769
step: 110, loss: 0.01166080404073
step: 120, loss: 0.061626631766557693
step: 130, loss: 0.014478799887001514

```

```

step: 140, loss: 0.01487412117421627
step: 150, loss: 0.011382977478206158
step: 160, loss: 0.0025711527559906244
step: 170, loss: 0.00675432151183486
step: 180, loss: 0.08634303510189056
step: 190, loss: 0.01410974096506834
step: 200, loss: 0.04067862033843994
step: 210, loss: 0.02993573248386383
step: 220, loss: 0.014182006940245628
step: 230, loss: 0.002218850888311863
step: 240, loss: 0.0031336310785263777
step: 250, loss: 0.014205480925738811
step: 260, loss: 0.0009218888008035719
step: 270, loss: 0.001536040217615664
step: 280, loss: 0.009480511769652367
step: 290, loss: 0.013226451352238655
step: 300, loss: 0.12731166183948517
step: 310, loss: 0.13701102137565613
step: 320, loss: 0.13797082006931305
step: 330, loss: 0.0037451668176800013
step: 340, loss: 0.0013523440575227141
step: 350, loss: 0.02319975383579731
step: 360, loss: 0.01642458140850067
step: 370, loss: 0.00848303735256195
step: 380, loss: 0.05959458649158478
step: 390, loss: 0.03531823307275772
step: 400, loss: 0.06879216432571411
step: 410, loss: 0.08728087693452835
step: 420, loss: 0.03280897065997124
step: 430, loss: 0.018394501879811287
step: 440, loss: 0.07625062763690948
step: 450, loss: 0.1215670108795166
step: 460, loss: 0.09616005420684814
step: 470, loss: 0.03152165189385414
step: 480, loss: 0.00967686902731657
step: 490, loss: 0.0019391807727515697
step: 500, loss: 0.11839286237955093
step: 510, loss: 0.00323274964466691
step: 520, loss: 0.007392133120447397
step: 530, loss: 0.05651891231536865
/usr/local/lib/python3.7/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=0.9238578680203047, f1=0.9238532110091744,
best_f1=0.9242144177449169
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```


step: 0, loss: 0.007093409076333046
step: 10, loss: 0.05970223993062973
step: 20, loss: 0.0733371302485466
step: 30, loss: 0.005382689647376537
step: 40, loss: 0.0067461892031133175
step: 50, loss: 0.003317941213026643
step: 60, loss: 0.004367745015770197
step: 70, loss: 0.005116622429341078
step: 80, loss: 0.1056504026055336
step: 90, loss: 0.009249762631952763
step: 100, loss: 0.024514272809028625
step: 110, loss: 0.0006076670833863318
step: 120, loss: 0.002153016161173582
step: 130, loss: 0.008692211471498013
step: 140, loss: 0.007604093756526709
step: 150, loss: 0.0017208597855642438
step: 160, loss: 0.00650087371468544
step: 170, loss: 0.005715436302125454
step: 180, loss: 0.012988908216357231
step: 190, loss: 0.0021242955699563026
step: 200, loss: 0.005988141521811485
step: 210, loss: 0.07059955596923828
step: 220, loss: 0.003743609879165888
step: 230, loss: 0.23753298819065094
step: 240, loss: 0.0029964016284793615
step: 250, loss: 0.0026562740094959736
step: 260, loss: 0.08939051628112793
step: 270, loss: 0.0359862744808197
step: 280, loss: 0.010378849692642689
step: 290, loss: 0.014996722340583801
step: 300, loss: 0.003573555964976549
step: 310, loss: 0.14255370199680328
step: 320, loss: 0.02867809124290943
step: 330, loss: 0.13687029480934143
step: 340, loss: 0.017359938472509384
step: 350, loss: 0.0012382343411445618
step: 360, loss: 0.03964199125766754
step: 370, loss: 0.11279743164777756
step: 380, loss: 0.002520510461181402
step: 390, loss: 0.04575127363204956
step: 400, loss: 0.029468750581145287
step: 410, loss: 0.09151281416416168
step: 420, loss: 0.08219612389802933
step: 430, loss: 0.00890355184674263
step: 440, loss: 0.024682123214006424
step: 450, loss: 0.005467654205858707
step: 460, loss: 0.002394520677626133
step: 470, loss: 0.0034170933067798615

```

step: 480, loss: 0.0016716046957299113
step: 490, loss: 0.09171529859304428
step: 500, loss: 0.012732299044728279
step: 510, loss: 0.15225279331207275
step: 520, loss: 0.014935131184756756
step: 530, loss: 0.0024875816889107227
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.932274638019617, f1=0.9167828916782892,
best_f1=0.9167828916782892
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0018858733819797635
step: 10, loss: 0.02907906472682953
step: 20, loss: 0.0018533134134486318
step: 30, loss: 0.0012092151446267962
step: 40, loss: 0.011182799935340881
step: 50, loss: 0.004671954549849033
step: 60, loss: 0.0007486275280825794
step: 70, loss: 0.0007764273905195296
step: 80, loss: 0.0024711296427994967
step: 90, loss: 0.0017202665330842137
step: 100, loss: 0.0005415093037299812
step: 110, loss: 0.00449305260553956
step: 120, loss: 0.01520059909671545
step: 130, loss: 0.003987722098827362
step: 140, loss: 0.019074169918894768
step: 150, loss: 0.0007871239795349538
step: 160, loss: 0.26170051097869873
step: 170, loss: 0.14707686007022858
step: 180, loss: 0.001094240928068757
step: 190, loss: 0.0007158999796956778
step: 200, loss: 0.058235056698322296
step: 210, loss: 0.0018260773504152894
step: 220, loss: 0.0004564083064906299
step: 230, loss: 0.002658218378201127
step: 240, loss: 0.010869953781366348
step: 250, loss: 0.000994111760519445
step: 260, loss: 0.005257365759462118
step: 270, loss: 0.003367051715031266
step: 280, loss: 0.04005013033747673
step: 290, loss: 0.000429338775575161

```

```

step: 300, loss: 0.03219733014702797
step: 310, loss: 0.001817976008169353
step: 320, loss: 0.002690332941710949
step: 330, loss: 0.0033722463995218277
step: 340, loss: 0.0008561449940316379
step: 350, loss: 0.00585330231115222
step: 360, loss: 0.0027038133703172207
step: 370, loss: 0.009693609550595284
step: 380, loss: 0.014720925129950047
step: 390, loss: 0.0011209809454157948
step: 400, loss: 0.004108942113816738
step: 410, loss: 0.0005670379032380879
step: 420, loss: 0.0009704145486466587
step: 430, loss: 0.006070210598409176
step: 440, loss: 0.022844135761260986
step: 450, loss: 0.00501785846427083
step: 460, loss: 0.013661406002938747
step: 470, loss: 0.004388580098748207
step: 480, loss: 0.0014532859204337
step: 490, loss: 0.002157962881028652
step: 500, loss: 0.019892888143658638
step: 510, loss: 0.001493746996857226
step: 520, loss: 0.002397012896835804
step: 530, loss: 0.002888440154492855
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9354389224338133, f1=0.9228611500701264,
best_f1=0.9228611500701264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005722080823034048
step: 10, loss: 0.0008622393361292779
step: 20, loss: 0.00045612381654791534
step: 30, loss: 0.011789781041443348
step: 40, loss: 0.0004660187987610698
step: 50, loss: 0.0017683359328657389
step: 60, loss: 0.0008976061944849789
step: 70, loss: 0.14625437557697296
step: 80, loss: 0.0004293580132070929
step: 90, loss: 0.03384605422616005
step: 100, loss: 0.18850889801979065
step: 110, loss: 0.003356640227138996

```

```
step: 120, loss: 0.0051264688372612
step: 130, loss: 0.00043034233385697007
step: 140, loss: 0.057531338185071945
step: 150, loss: 0.016214454546570778
step: 160, loss: 0.019232086837291718
step: 170, loss: 0.001533397939056158
step: 180, loss: 0.00024238333571702242
step: 190, loss: 0.0013453580904752016
step: 200, loss: 0.00043633655877783895
step: 210, loss: 0.0005420061643235385
step: 220, loss: 0.00265116011723876
step: 230, loss: 0.00022380199516192079
step: 240, loss: 0.007289353292435408
step: 250, loss: 0.002143088961020112
step: 260, loss: 0.00048699890612624586
step: 270, loss: 0.002144854050129652
step: 280, loss: 0.0005744410445913672
step: 290, loss: 0.016981948167085648
step: 300, loss: 0.0007707082550041378
step: 310, loss: 0.0009835273958742619
step: 320, loss: 0.00271313707344234
step: 330, loss: 0.0007005502702668309
step: 340, loss: 0.0012594081927090883
step: 350, loss: 0.023647090420126915
step: 360, loss: 0.009394880384206772
step: 370, loss: 0.009440219029784203
step: 380, loss: 0.02096232771873474
step: 390, loss: 0.05342389643192291
step: 400, loss: 0.0002923432330135256
step: 410, loss: 0.03521415963768959
step: 420, loss: 0.002406845800578594
step: 430, loss: 0.0024867216125130653
step: 440, loss: 0.05403765290975571
step: 450, loss: 0.0046358294785022736
step: 460, loss: 0.05636953189969063
step: 470, loss: 0.1658691018819809
step: 480, loss: 0.008707213215529919
step: 490, loss: 0.001558555057272315
step: 500, loss: 0.0021068123169243336
step: 510, loss: 0.00036257036845199764
step: 520, loss: 0.03252921625971794
step: 530, loss: 0.020576391369104385
/usr/local/lib/python3.7/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=0.9202453987730062, f1=0.9144208037825059,
best_f1=0.9228611500701264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

An input tensor was not cuda.

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

```
step: 0, loss: 0.004014987964183092
step: 10, loss: 0.01507976371794939
step: 20, loss: 0.02694319561123848
step: 30, loss: 0.0025066453963518143
step: 40, loss: 0.0009595154551789165
step: 50, loss: 0.001603846438229084
step: 60, loss: 0.0035276280250400305
step: 70, loss: 0.0006324120331555605
step: 80, loss: 0.0032835975289344788
step: 90, loss: 0.0005251033580861986
step: 100, loss: 0.000449798273621127
step: 110, loss: 0.0004117197822779417
step: 120, loss: 0.00011376863403711468
step: 130, loss: 0.0010326869087293744
step: 140, loss: 0.0002830212761182338
step: 150, loss: 0.00016812185640446842
step: 160, loss: 0.0006213756860233843
step: 170, loss: 0.03651094436645508
step: 180, loss: 0.00406988151371479
step: 190, loss: 0.00028358626877889037
step: 200, loss: 0.006471225060522556
step: 210, loss: 0.0029529626481235027
step: 220, loss: 0.002243290888145566
step: 230, loss: 0.0014409208670258522
step: 240, loss: 0.007826454006135464
step: 250, loss: 0.000787412456702441
step: 260, loss: 0.0012522975448518991
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 270, loss: 0.00016588589642196894
step: 280, loss: 0.22298923134803772
step: 290, loss: 0.010991917923092842
step: 300, loss: 0.00019070581765845418
step: 310, loss: 0.0012902435846626759
step: 320, loss: 0.03896210342645645
step: 330, loss: 0.00020795695309061557
step: 340, loss: 0.012843331322073936
step: 350, loss: 0.0004121562233194709
step: 360, loss: 0.0009888604981824756
step: 370, loss: 0.0005420089582912624
step: 380, loss: 0.0028950090054422617
step: 390, loss: 0.01889132335782051
step: 400, loss: 0.0068964590318500996
step: 410, loss: 0.021858014166355133
step: 420, loss: 0.0003499901795294136
step: 430, loss: 9.12671530386433e-05
step: 440, loss: 0.001891481108032167
```

```

step: 450, loss: 0.000767870747949928
step: 460, loss: 0.002117869211360812
step: 470, loss: 0.004850597586482763
step: 480, loss: 0.018274201080203056
step: 490, loss: 0.001788100227713585
step: 500, loss: 0.03139638155698776
step: 510, loss: 0.0011524385772645473
step: 520, loss: 0.0005275901639834046
step: 530, loss: 0.003581464057788253
/usr/local/lib/python3.7/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=0.9255663430420712, f1=0.9256880733944954,
best_f1=0.9228611500701264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0017740594921633601
step: 10, loss: 0.0013209491735324264
step: 20, loss: 0.0003182673826813698
step: 30, loss: 0.000707075116224587
step: 40, loss: 0.00078899587970227
step: 50, loss: 0.00039763981476426125
step: 60, loss: 0.00013918588228989393
step: 70, loss: 0.0005061220144852996
step: 80, loss: 0.00012567441444844007
step: 90, loss: 6.880379078211263e-05
step: 100, loss: 0.0024198065511882305
step: 110, loss: 0.001985592069104314
step: 120, loss: 0.0005561466095969081
step: 130, loss: 0.0007341146119870245
step: 140, loss: 6.0326914535835385e-05
step: 150, loss: 0.005174386315047741
step: 160, loss: 0.00020753109129145741
step: 170, loss: 0.05949923396110535
step: 180, loss: 0.00045331398723647
step: 190, loss: 0.000915033626370132
step: 200, loss: 0.0015692032175138593
step: 210, loss: 0.0022729046177119017
step: 220, loss: 0.0018188490066677332
step: 230, loss: 0.00047234425437636673
step: 240, loss: 9.028327622218058e-05
step: 250, loss: 0.003420345950871706
step: 260, loss: 0.03243149444460869
step: 270, loss: 8.316936873598024e-05
step: 280, loss: 0.0020407559350132942
step: 290, loss: 0.0007107815472409129
step: 300, loss: 0.0026304125785827637

```

```

step: 310, loss: 0.03540579602122307
step: 320, loss: 0.0003191927680745721
step: 330, loss: 0.03872235491871834
step: 340, loss: 0.0007831083494238555
step: 350, loss: 0.06234653666615486
step: 360, loss: 0.00031363131711259484
step: 370, loss: 0.0007113652536645532
step: 380, loss: 0.0008050660835579038
step: 390, loss: 0.0002330811257706955
step: 400, loss: 0.2130606472492218
step: 410, loss: 0.00012630071432795376
step: 420, loss: 0.02998119592666626
step: 430, loss: 0.00011768681724788621
step: 440, loss: 0.0002056246012216434
step: 450, loss: 0.0010862129274755716
step: 460, loss: 0.0003288611478637904
step: 470, loss: 0.00035867016413249075
step: 480, loss: 0.00023246144701261073
step: 490, loss: 0.0006725783459842205
step: 500, loss: 0.000522975402418524
step: 510, loss: 0.0002832047757692635
step: 520, loss: 0.00013210887846071273
step: 530, loss: 0.0002546102914493531
/usr/local/lib/python3.7/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=0.9263059701492538, f1=0.9208566108007449,
best_f1=0.9228611500701264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00010265397577313706
step: 10, loss: 0.004690211266279221
step: 20, loss: 0.0008160985307767987
step: 30, loss: 0.0013612181646749377
step: 40, loss: 0.0037268579471856356
step: 50, loss: 0.00030930506181903183
step: 60, loss: 0.0005579768912866712
step: 70, loss: 0.10427548736333847
step: 80, loss: 0.0004886702517978847
step: 90, loss: 0.0004976914497092366
step: 100, loss: 0.0007566220010630786
step: 110, loss: 0.0025997215416282415
step: 120, loss: 0.0007657061796635389
step: 130, loss: 0.0006859512068331242
step: 140, loss: 0.0014613483799621463
step: 150, loss: 0.00034528604010120034
step: 160, loss: 0.0010826083598658442

```

```

step: 170, loss: 0.00013516846229322255
step: 180, loss: 0.005000299774110317
step: 190, loss: 0.0006531914696097374
step: 200, loss: 0.0005971417995169759
step: 210, loss: 8.973589865490794e-05
step: 220, loss: 0.0013359417207539082
step: 230, loss: 0.0004908725386485457
step: 240, loss: 0.004960689228028059
step: 250, loss: 8.873204205883667e-05
step: 260, loss: 0.004417323041707277
step: 270, loss: 0.004517655819654465
step: 280, loss: 0.000809239165391773
step: 290, loss: 8.761257049627602e-05
step: 300, loss: 0.00040000397711992264
step: 310, loss: 0.0006948571535758674
step: 320, loss: 0.014412241987884045
step: 330, loss: 5.89747323829215e-05
step: 340, loss: 4.33036511822138e-05
step: 350, loss: 5.863586557097733e-05
step: 360, loss: 0.013437826186418533
step: 370, loss: 0.00013332664093468338
step: 380, loss: 0.0006366941379383206
step: 390, loss: 0.00015452438674401492
step: 400, loss: 0.002030491130426526
step: 410, loss: 0.00043228172580711544
step: 420, loss: 0.03501339256763458
step: 430, loss: 5.519818660104647e-05
step: 440, loss: 9.909797518048435e-05
step: 450, loss: 5.194275581743568e-05
step: 460, loss: 0.0001769232767401263
step: 470, loss: 0.00011806697148131207
step: 480, loss: 0.0007777712889946997
step: 490, loss: 0.00013397441944107413
step: 500, loss: 0.018429415300488472
step: 510, loss: 0.003388607408851385
step: 520, loss: 7.669108163099736e-05
step: 530, loss: 7.816453580744565e-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=0.9272559852670351, f1=0.9178654292343388,
best_f1=0.9228611500701264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00016566987324040383
step: 10, loss: 4.882661232841201e-05
step: 20, loss: 0.00029326966614462435

```


step: 30, loss: 0.0002130977518390864
step: 40, loss: 0.000890035298652947
step: 50, loss: 0.0004310337535571307
step: 60, loss: 0.00014972286589909345
step: 70, loss: 0.0006159866461530328
step: 80, loss: 0.0006808809703215957
step: 90, loss: 0.000710902560967952
step: 100, loss: 0.0008853487088344991
step: 110, loss: 0.001968187279999256
step: 120, loss: 0.00021851653582416475
step: 130, loss: 6.13485899521038e-05
step: 140, loss: 0.0011085344012826681
step: 150, loss: 0.00029128577443771064
step: 160, loss: 8.404208347201347e-05
step: 170, loss: 7.966706471052021e-05
step: 180, loss: 0.03355761617422104
step: 190, loss: 0.006051075179129839
step: 200, loss: 0.00599158089607954
step: 210, loss: 8.755733870202675e-05
step: 220, loss: 0.03845294564962387
step: 230, loss: 0.0010402792831882834
step: 240, loss: 4.7814377467148006e-05
step: 250, loss: 7.036274473648518e-05
step: 260, loss: 0.00022260809782892466
step: 270, loss: 0.00599170196801424
step: 280, loss: 7.493576413253322e-05
step: 290, loss: 0.0001008419058052823
step: 300, loss: 0.0015608955873176455
step: 310, loss: 8.319292828673497e-05
step: 320, loss: 3.1447114452021196e-05
step: 330, loss: 0.00020791245333384722
step: 340, loss: 4.2848729208344594e-05
step: 350, loss: 3.998929969384335e-05
step: 360, loss: 0.00865771435201168
step: 370, loss: 0.0007845880463719368
step: 380, loss: 0.00029024324612692
step: 390, loss: 0.000557350052986294
step: 400, loss: 0.00022341353178489953
step: 410, loss: 0.0016227560117840767
step: 420, loss: 0.026265546679496765
step: 430, loss: 0.0018175262957811356
step: 440, loss: 0.00037368517951108515
step: 450, loss: 0.00020979622786398977
step: 460, loss: 0.0018325423588976264
step: 470, loss: 7.195998477982357e-05
step: 480, loss: 0.007743007969111204
step: 490, loss: 0.04508811607956886
step: 500, loss: 0.002929914044216275

```

step: 510, loss: 0.0005432405159808695
step: 520, loss: 0.00014256844588089734
step: 530, loss: 0.0017885337583720684
/usr/local/lib/python3.7/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=0.9241706161137442, f1=0.9170638703527169,
best_f1=0.9228611500701264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002282516099512577
step: 10, loss: 0.0021801271941512823
step: 20, loss: 0.0016111828153952956
step: 30, loss: 0.00030762769165448844
step: 40, loss: 0.0029256027191877365
step: 50, loss: 0.002524858806282282
step: 60, loss: 0.00013878375466447324
step: 70, loss: 0.00012660666834563017
step: 80, loss: 0.00013308523921296
step: 90, loss: 0.00017811550060287118
step: 100, loss: 0.0002444958663545549
step: 110, loss: 0.00016731504001654685
step: 120, loss: 6.345351721392944e-05
step: 130, loss: 8.51561562740244e-05
step: 140, loss: 6.593984289793298e-05
step: 150, loss: 0.0006682446110062301
step: 160, loss: 0.006280418951064348
step: 170, loss: 0.001234617317095399
step: 180, loss: 4.508743586484343e-05
step: 190, loss: 0.019357893615961075
step: 200, loss: 0.00022136095503810793
step: 210, loss: 7.273542723851278e-05
step: 220, loss: 9.597935422789305e-05
step: 230, loss: 3.9589693187735975e-05
step: 240, loss: 5.059270915808156e-05
step: 250, loss: 6.668100832030177e-05
step: 260, loss: 0.0001420355256414041
step: 270, loss: 0.0034395279362797737
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 280, loss: 0.0008389076683670282
step: 290, loss: 0.0009162828209809959
step: 300, loss: 0.05711226910352707
step: 310, loss: 0.009223508648574352
step: 320, loss: 0.00269345217384398
step: 330, loss: 0.002521777991205454
step: 340, loss: 0.022801563143730164
step: 350, loss: 0.0007701707654632628

```

```

step: 360, loss: 0.000900350627489388
step: 370, loss: 3.0124716431600973e-05
step: 380, loss: 3.032789027201943e-05
step: 390, loss: 0.0018913047388195992
step: 400, loss: 4.484668170334771e-05
step: 410, loss: 0.0006516897701658309
step: 420, loss: 0.0004426347150001675
step: 430, loss: 0.0027906899340450764
step: 440, loss: 8.213089313358068e-05
step: 450, loss: 3.3637843444012105e-05
step: 460, loss: 0.000626339518930763
step: 470, loss: 0.0013315420364961028
step: 480, loss: 5.890888860449195e-05
step: 490, loss: 0.0010623878333717585
step: 500, loss: 0.0004760913725476712
step: 510, loss: 3.807768371189013e-05
step: 520, loss: 3.3797798096202314e-05
step: 530, loss: 2.068219328066334e-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=0.9203624225083452, f1=0.9122137404580153,
best_f1=0.9228611500701264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001436591992387548
step: 10, loss: 0.0004372672119643539
step: 20, loss: 6.649043643847108e-05
step: 30, loss: 0.00027795028290711343
step: 40, loss: 4.095999975106679e-05
step: 50, loss: 0.00019073639123234898
step: 60, loss: 3.9780079532647505e-05
step: 70, loss: 6.34667812846601e-05
step: 80, loss: 0.0008877580985426903
step: 90, loss: 9.509595111012459e-05
step: 100, loss: 5.396501364884898e-05
step: 110, loss: 3.254576586186886e-05
step: 120, loss: 4.8882910050451756e-05
step: 130, loss: 5.182551103644073e-05
step: 140, loss: 9.459999273531139e-05
step: 150, loss: 0.0013868536334484816
step: 160, loss: 3.339954491821118e-05
step: 170, loss: 5.654554843204096e-05
step: 180, loss: 0.005027859006077051
step: 190, loss: 0.002529645338654518
step: 200, loss: 3.092876431765035e-05
step: 210, loss: 0.006329790689051151

```

```

step: 220, loss: 2.7320369554217905e-05
step: 230, loss: 0.0014726904919371009
step: 240, loss: 5.9893576690228656e-05
step: 250, loss: 2.7402124032960273e-05
step: 260, loss: 0.0010477026225998998
step: 270, loss: 2.758066329988651e-05
step: 280, loss: 3.761170955840498e-05
step: 290, loss: 9.159419278148562e-05
step: 300, loss: 6.335224315989763e-05
step: 310, loss: 4.2530147766228765e-05
step: 320, loss: 0.0001029886188916862
step: 330, loss: 2.9775277653243393e-05
step: 340, loss: 0.009416729211807251
step: 350, loss: 0.0015604164218530059
step: 360, loss: 0.00042483254219405353
step: 370, loss: 4.715679824585095e-05
step: 380, loss: 3.985775401815772e-05
step: 390, loss: 4.8184847400989383e-05
step: 400, loss: 3.7403082387754694e-05
step: 410, loss: 0.0443761833012104
step: 420, loss: 0.00045956563553772867
step: 430, loss: 4.37947383034043e-05
step: 440, loss: 9.917841089190915e-05
step: 450, loss: 7.586546416860074e-05
step: 460, loss: 0.007913921028375626
step: 470, loss: 3.202066363883205e-05
step: 480, loss: 8.837974019115791e-05
step: 490, loss: 3.575268419808708e-05
step: 500, loss: 0.0018282290548086166
step: 510, loss: 5.7665787608129904e-05
step: 520, loss: 0.0003336795780342072
step: 530, loss: 0.0010315357940271497
/usr/local/lib/python3.7/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=0.9277389277389277, f1=0.9159663865546218,
best_f1=0.9228611500701264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001468669215682894
step: 10, loss: 3.751191252376884e-05
step: 20, loss: 0.009793858975172043
step: 30, loss: 0.025747274979948997
step: 40, loss: 5.106506068841554e-05
step: 50, loss: 0.0001564793783472851
step: 60, loss: 3.1171693990472704e-05
step: 70, loss: 3.909203951479867e-05

```

step: 80, loss: 2.037317244685255e-05
step: 90, loss: 8.072410128079355e-05
step: 100, loss: 2.521517308196053e-05
step: 110, loss: 0.022446315735578537
step: 120, loss: 5.680831236531958e-05
step: 130, loss: 0.00034418850555084646
step: 140, loss: 6.303460395429283e-05
step: 150, loss: 0.00015989977691788226
step: 160, loss: 0.000150654639583081
step: 170, loss: 0.0005803022067993879
step: 180, loss: 2.3673501345911063e-05
step: 190, loss: 4.4276042899582535e-05
step: 200, loss: 0.0031159676145762205
step: 210, loss: 0.0028086663223803043
step: 220, loss: 0.05383564531803131
step: 230, loss: 3.151009514112957e-05
step: 240, loss: 0.000186327044502832
step: 250, loss: 0.0037186455447226763
step: 260, loss: 2.5748095140443183e-05
step: 270, loss: 0.00011533693759702146
step: 280, loss: 5.201882959227078e-05
step: 290, loss: 5.341309224604629e-05
step: 300, loss: 0.0006080453749746084
step: 310, loss: 0.0013313292292878032
step: 320, loss: 5.287718158797361e-05
step: 330, loss: 9.974334534490481e-05
step: 340, loss: 2.7149142624693923e-05
step: 350, loss: 0.10731294751167297
step: 360, loss: 3.9781181840226054e-05
step: 370, loss: 3.0999570299172774e-05
step: 380, loss: 0.000325435510603711
step: 390, loss: 6.855756510049105e-05
step: 400, loss: 2.8466583898989484e-05
step: 410, loss: 0.0007710197824053466
step: 420, loss: 0.00015610815898980945
step: 430, loss: 5.6235341617139056e-05
step: 440, loss: 6.665611726930365e-05
step: 450, loss: 0.00012073892139596865
step: 460, loss: 2.214244523202069e-05
step: 470, loss: 0.009351559914648533
step: 480, loss: 0.0002364974352531135
step: 490, loss: 0.00013482112262863666
step: 500, loss: 0.05520373582839966
step: 510, loss: 3.431186269153841e-05
step: 520, loss: 6.179117917781696e-05
step: 530, loss: 0.00013082570512779057
/usr/local/lib/python3.7/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=0.9263256687001408, f1=0.9129411764705881,
best_f1=0.9228611500701264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.204437522683293e-05
step: 10, loss: 0.003517615143209696
step: 20, loss: 0.00028185255359858274
step: 30, loss: 9.448142373003066e-05
step: 40, loss: 0.0003335486399009824
step: 50, loss: 7.533900497946888e-05
step: 60, loss: 4.4803036871599033e-05
step: 70, loss: 3.5249766369815916e-05
step: 80, loss: 3.613705121097155e-05
step: 90, loss: 2.3487062208005227e-05
step: 100, loss: 0.00035258286516182125
step: 110, loss: 3.373909930814989e-05
step: 120, loss: 0.00010227266466245055
step: 130, loss: 3.465603367658332e-05
step: 140, loss: 0.0033643078058958054
step: 150, loss: 4.7108682338148355e-05
step: 160, loss: 0.0013056329917162657
step: 170, loss: 0.00025325786555185914
step: 180, loss: 3.494851625873707e-05
step: 190, loss: 0.0003575546434149146
step: 200, loss: 5.534854426514357e-05
step: 210, loss: 0.0001074518368113786
step: 220, loss: 2.180347109970171e-05
step: 230, loss: 0.00024129453231580555
step: 240, loss: 0.0027516186237335205
step: 250, loss: 3.44193285854999e-05
step: 260, loss: 1.7404268874088302e-05
step: 270, loss: 0.001401966786943376
step: 280, loss: 0.0020673288963735104
step: 290, loss: 3.210989598301239e-05
step: 300, loss: 2.4277043848996982e-05
step: 310, loss: 3.369831028976478e-05
step: 320, loss: 0.00014985665620770305
step: 330, loss: 3.027574894076679e-05
step: 340, loss: 3.292609108029865e-05
step: 350, loss: 9.05203924048692e-05
step: 360, loss: 0.0013582572573795915
step: 370, loss: 3.41675549861975e-05
step: 380, loss: 3.191442010574974e-05
step: 390, loss: 2.298436447745189e-05
step: 400, loss: 2.8773332815035246e-05
step: 410, loss: 0.00010430450493004173
```

```

step: 420, loss: 1.818656164687127e-05
step: 430, loss: 2.8377597118378617e-05
step: 440, loss: 1.318735030508833e-05
step: 450, loss: 2.3513237465522252e-05
step: 460, loss: 4.4721062295138836e-05
step: 470, loss: 0.0004066166002303362
step: 480, loss: 6.52197704766877e-05
step: 490, loss: 0.0009113184059970081
step: 500, loss: 0.0644117146730423
step: 510, loss: 2.373679672018625e-05
step: 520, loss: 2.7808346203528345e-05
step: 530, loss: 2.3926811991259456e-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=0.9271274094969439, f1=0.9138012246820537,
best_f1=0.9228611500701264
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 2.9029144570813514e-05
step: 10, loss: 0.00010143604595214128
step: 20, loss: 8.693463314557448e-05
step: 30, loss: 5.020086609874852e-05
step: 40, loss: 2.417115683783777e-05
step: 50, loss: 4.8929279728326946e-05
step: 60, loss: 1.2621129826584365e-05
step: 70, loss: 6.56253396300599e-05
step: 80, loss: 5.1206458010710776e-05
step: 90, loss: 1.6048285033321008e-05
step: 100, loss: 1.665552736085374e-05
step: 110, loss: 0.0015973341651260853
step: 120, loss: 0.0007103780517354608
step: 130, loss: 1.7415386537322775e-05
step: 140, loss: 1.4412932614504825e-05
step: 150, loss: 0.0002651189861353487
step: 160, loss: 2.77072649623733e-05
step: 170, loss: 3.5946082789450884e-05
step: 180, loss: 5.470225369208492e-05
step: 190, loss: 4.1616789530962706e-05
step: 200, loss: 3.778708924073726e-05
step: 210, loss: 9.163887443719432e-05
step: 220, loss: 1.3798263353237417e-05
step: 230, loss: 8.016225910978392e-05
step: 240, loss: 4.144840204389766e-05
step: 250, loss: 7.025556988082826e-05
step: 260, loss: 0.0002306328242411837
step: 270, loss: 0.00020314680295996368

```

```

step: 280, loss: 3.0109877116046846e-05
step: 290, loss: 0.002182355849072337
step: 300, loss: 0.0013174820924177766
step: 310, loss: 0.00010877282329602167
step: 320, loss: 0.00016256835078820586
step: 330, loss: 0.00021145310893189162
step: 340, loss: 4.386056025396101e-05
step: 350, loss: 0.0003924092452507466
step: 360, loss: 3.021450902451761e-05
step: 370, loss: 1.856638482422568e-05
step: 380, loss: 3.306716462248005e-05
step: 390, loss: 1.9791803424595855e-05
step: 400, loss: 4.829233148484491e-05
step: 410, loss: 5.4252399422694e-05
step: 420, loss: 2.1982461476000026e-05
step: 430, loss: 2.1635878511006013e-05
step: 440, loss: 3.466443013167009e-05
step: 450, loss: 0.015410752035677433
step: 460, loss: 1.4822720004303847e-05
step: 470, loss: 2.9196735340519808e-05
step: 480, loss: 0.00025335626560263336
step: 490, loss: 0.00025870109675452113
step: 500, loss: 3.09030401695054e-05
step: 510, loss: 7.718663982814178e-05
step: 520, loss: 2.4824452339089476e-05
step: 530, loss: 2.13528037420474e-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=0.9263456090651558, f1=0.9123638086215065,
best_f1=0.9228611500701264
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

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

```



```

patch_torch_functions : False
keep_batchnorm_fp32   : True
master_weights        : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
5742it [00:14, 387.04it/s]
load_f1 = 0.9353647276084949
real_f1 = 0.9330275229357798
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, 415.28it/s]

```

1.1.4 Amazon-Google - Running the matcher

```

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

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

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

```

```

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

```

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

Defaults for this optimization level are:

```

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

```

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

After processing overrides, optimization options are:

```

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

```

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

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

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

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

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

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

step: 0, loss: 0.848127007484436

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.06646598875522614

```

step: 20, loss: 0.3751990497112274
step: 30, loss: 0.3723691403865814
step: 40, loss: 0.499857097864151
step: 50, loss: 0.29021841287612915
step: 60, loss: 0.3618883192539215
step: 70, loss: 0.22780069708824158
step: 80, loss: 0.2991541922092438
step: 90, loss: 0.40986722707748413
step: 100, loss: 0.17506273090839386
step: 110, loss: 0.32177767157554626
step: 120, loss: 0.24560658633708954
step: 130, loss: 0.23766207695007324
step: 140, loss: 0.23838229477405548
step: 150, loss: 0.24812383949756622
step: 160, loss: 0.21085795760154724
step: 170, loss: 0.11445928364992142
step: 180, loss: 0.1927461326122284
step: 190, loss: 0.31349435448646545
step: 200, loss: 0.12864474952220917
step: 210, loss: 0.3909052014350891
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5223613595706619, f1=0.5637065637065637,
best_f1=0.5637065637065637
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07611711323261261
step: 10, loss: 0.05479808896780014
step: 20, loss: 0.322761207818985
step: 30, loss: 0.1072038859128952
step: 40, loss: 0.09986897557973862
step: 50, loss: 0.19305481016635895
step: 60, loss: 0.038043856620788574
step: 70, loss: 0.20581424236297607
step: 80, loss: 0.1726676970720291
step: 90, loss: 0.13139505684375763
step: 100, loss: 0.0729328989982605
step: 110, loss: 0.11728842556476593
step: 120, loss: 0.13080398738384247
step: 130, loss: 0.20004625618457794
step: 140, loss: 0.24367459118366241
step: 150, loss: 0.14622685313224792

```

```

step: 160, loss: 0.09742506593465805
step: 170, loss: 0.19144928455352783
step: 180, loss: 0.2694200873374939
step: 190, loss: 0.29160669445991516
step: 200, loss: 0.18674524128437042
step: 210, loss: 0.2603233754634857
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6150870406189556, f1=0.6055776892430279,
best_f1=0.6055776892430279
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.17083962261676788
step: 10, loss: 0.2779208719730377
step: 20, loss: 0.2669926583766937
step: 30, loss: 0.09972096234560013
step: 40, loss: 0.12472455203533173
step: 50, loss: 0.10734181851148605
step: 60, loss: 0.1994042843580246
step: 70, loss: 0.1266535371541977
step: 80, loss: 0.0676644816994667
step: 90, loss: 0.045809242874383926
step: 100, loss: 0.01729593612253666
step: 110, loss: 0.294792503118515
step: 120, loss: 0.1877296268939972
step: 130, loss: 0.1913214474916458
step: 140, loss: 0.2756892144680023
step: 150, loss: 0.15891259908676147
step: 160, loss: 0.0585227906703949
step: 170, loss: 0.282620906829834
step: 180, loss: 0.15936152637004852
step: 190, loss: 0.18338380753993988
step: 200, loss: 0.14263056218624115
step: 210, loss: 0.12850281596183777
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6201232032854209, f1=0.6205450733752621,

```

```

best_f1=0.6205450733752621
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.12460686266422272
step: 10, loss: 0.04694842919707298
step: 20, loss: 0.12860779464244843
step: 30, loss: 0.01886197179555893
step: 40, loss: 0.04344618692994118
step: 50, loss: 0.03914228826761246
step: 60, loss: 0.042852938175201416
step: 70, loss: 0.1167074590921402
step: 80, loss: 0.2044830322265625
step: 90, loss: 0.006225674878805876
step: 100, loss: 0.05998130142688751
step: 110, loss: 0.18912726640701294
step: 120, loss: 0.11471236497163773
step: 130, loss: 0.1089981347322464
step: 140, loss: 0.20920813083648682
step: 150, loss: 0.0914573147892952
step: 160, loss: 0.04951278120279312
step: 170, loss: 0.060306791216135025
step: 180, loss: 0.16467292606830597
step: 190, loss: 0.05760795623064041
step: 200, loss: 0.08565434068441391
step: 210, loss: 0.04418722540140152
/usr/local/lib/python3.7/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=0.5962962962962962, f1=0.6309963099630995,
best_f1=0.6205450733752621
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1164654940366745
step: 10, loss: 0.1360408514738083
step: 20, loss: 0.055788975208997726
step: 30, loss: 0.09348088502883911
step: 40, loss: 0.2044127732515335
step: 50, loss: 0.11875695735216141
step: 60, loss: 0.1163407638669014
step: 70, loss: 0.07827599346637726
step: 80, loss: 0.08064067363739014
step: 90, loss: 0.18088699877262115
step: 100, loss: 0.04656991735100746
step: 110, loss: 0.01685636304318905
step: 120, loss: 0.016085006296634674
step: 130, loss: 0.07063356041908264

```

```

step: 140, loss: 0.14542049169540405
step: 150, loss: 0.04713236540555954
step: 160, loss: 0.010755148716270924
step: 170, loss: 0.17017658054828644
step: 180, loss: 0.11642421036958694
step: 190, loss: 0.11933279782533646
step: 200, loss: 0.022012554109096527
step: 210, loss: 0.027444856241345406
/usr/local/lib/python3.7/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=0.6166666666666667, f1=0.6201232032854209,
best_f1=0.6205450733752621
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.059508319944143295
step: 10, loss: 0.037320446223020554
step: 20, loss: 0.015429294668138027
step: 30, loss: 0.16321289539337158
step: 40, loss: 0.04132715240120888
step: 50, loss: 0.042103566229343414
step: 60, loss: 0.1662137806415558
step: 70, loss: 0.13694611191749573
step: 80, loss: 0.07564737647771835
step: 90, loss: 0.09931435436010361
step: 100, loss: 0.017703723162412643
step: 110, loss: 0.08453426510095596
step: 120, loss: 0.056937575340270996
step: 130, loss: 0.06457717716693878
step: 140, loss: 0.03416704386472702
step: 150, loss: 0.04088640585541725
step: 160, loss: 0.19483010470867157
step: 170, loss: 0.09782277792692184
step: 180, loss: 0.0057178279384970665
step: 190, loss: 0.10779768228530884
step: 200, loss: 0.010322343558073044
step: 210, loss: 0.10390754789113998
/usr/local/lib/python3.7/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=0.6176470588235294, f1=0.6045548654244307,
best_f1=0.6205450733752621
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04390519857406616
step: 10, loss: 0.07673151791095734

```

```

step: 20, loss: 0.08910143375396729
step: 30, loss: 0.004396913573145866
step: 40, loss: 0.03847670182585716
step: 50, loss: 0.07626447826623917
step: 60, loss: 0.0862041562795639
step: 70, loss: 0.006161891855299473
step: 80, loss: 0.07624294608831406
step: 90, loss: 0.026769505813717842
step: 100, loss: 0.01142976712435484
step: 110, loss: 0.24424920976161957
step: 120, loss: 0.15054024755954742
step: 130, loss: 0.03212787210941315
step: 140, loss: 0.0044175428338348866
step: 150, loss: 0.02908196859061718
step: 160, loss: 0.015677422285079956
step: 170, loss: 0.019679874181747437
step: 180, loss: 0.00854149553924799
step: 190, loss: 0.031641192734241486
step: 200, loss: 0.00648488337174058
step: 210, loss: 0.051283594220876694
/usr/local/lib/python3.7/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=0.5991735537190083, f1=0.6206896551724138,
best_f1=0.6205450733752621
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.045233529061079025
step: 10, loss: 0.02223494090139866
step: 20, loss: 0.004630761221051216
step: 30, loss: 0.023850446566939354
step: 40, loss: 0.07896227389574051
step: 50, loss: 0.05108126252889633
step: 60, loss: 0.2427891492843628
step: 70, loss: 0.015910841524600983
step: 80, loss: 0.03466047719120979
step: 90, loss: 0.02541431598365307
step: 100, loss: 0.020222511142492294
step: 110, loss: 0.07131164520978928
step: 120, loss: 0.07341352105140686
step: 130, loss: 0.010203791782259941
step: 140, loss: 0.0020812228322029114
step: 150, loss: 0.005748506169766188
step: 160, loss: 0.034504324197769165
step: 170, loss: 0.02408476732671261
step: 180, loss: 0.04941651597619057
step: 190, loss: 0.05935528874397278

```

```

step: 200, loss: 0.1681118905544281
step: 210, loss: 0.1456964612007141
/usr/local/lib/python3.7/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=0.6063157894736841, f1=0.6053169734151329,
best_f1=0.6205450733752621
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.015567008405923843
step: 10, loss: 0.43317681550979614
step: 20, loss: 0.044087015092372894
step: 30, loss: 0.13024571537971497
step: 40, loss: 0.0757490023970604
step: 50, loss: 0.09595254063606262
step: 60, loss: 0.024861840531229973
step: 70, loss: 0.028746223077178
step: 80, loss: 0.24871979653835297
step: 90, loss: 0.12855617702007294
step: 100, loss: 0.03264676406979561
step: 110, loss: 0.025098685175180435
step: 120, loss: 0.01167142204940319
step: 130, loss: 0.031457070261240005
step: 140, loss: 0.0484439954161644
step: 150, loss: 0.011948184110224247
step: 160, loss: 0.002652991097420454
step: 170, loss: 0.15687738358974457
step: 180, loss: 0.03829187527298927
step: 190, loss: 0.18635182082653046
step: 200, loss: 0.04823767766356468
step: 210, loss: 0.10404720902442932
/usr/local/lib/python3.7/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=0.6032388663967612, f1=0.6085192697768762,
best_f1=0.6205450733752621
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.018939366564154625
step: 10, loss: 0.05572609230875969
step: 20, loss: 0.021900847554206848
step: 30, loss: 0.004035111516714096
step: 40, loss: 0.030333174392580986
step: 50, loss: 0.0031339649576693773
step: 60, loss: 0.0025470901746302843
step: 70, loss: 0.0018102079629898071

```



```

step: 80, loss: 0.0013957382179796696
step: 90, loss: 0.057858146727085114
step: 100, loss: 0.002541695721447468
step: 110, loss: 0.014673472382128239
step: 120, loss: 0.05910585820674896
step: 130, loss: 0.006566149182617664
step: 140, loss: 0.011145559139549732
step: 150, loss: 0.05242455005645752
step: 160, loss: 0.05420839041471481
step: 170, loss: 0.09445411711931229
step: 180, loss: 0.02501148171722889
step: 190, loss: 0.15504074096679688
step: 200, loss: 0.08116792142391205
step: 210, loss: 0.20559965074062347
/usr/local/lib/python3.7/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=0.5976095617529881, f1=0.6102362204724409,
best_f1=0.6205450733752621
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006362822838127613
step: 10, loss: 0.016646282747387886
step: 20, loss: 0.004283770453184843
step: 30, loss: 0.022232532501220703
step: 40, loss: 0.005760678555816412
step: 50, loss: 0.009147332981228828
step: 60, loss: 0.009914400056004524
step: 70, loss: 0.0014326535165309906
step: 80, loss: 0.2545279562473297
step: 90, loss: 0.12315733730792999
step: 100, loss: 0.05694270879030228
step: 110, loss: 0.0034830246586352587
step: 120, loss: 0.010655109770596027
step: 130, loss: 0.08389289677143097
step: 140, loss: 0.014512933790683746
step: 150, loss: 0.05655793845653534
step: 160, loss: 0.012175317853689194
step: 170, loss: 0.04453086853027344
step: 180, loss: 0.07974187284708023
step: 190, loss: 0.02491777203977108
step: 200, loss: 0.003020285628736019
step: 210, loss: 0.04269711300730705
/usr/local/lib/python3.7/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=0.6033755274261604, f1=0.6182572614107884,

```

```

best_f1=0.6205450733752621
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009435954503715038
step: 10, loss: 0.04166276380419731
step: 20, loss: 0.01521597895771265
step: 30, loss: 0.012099631130695343
step: 40, loss: 0.06377792358398438
step: 50, loss: 0.03411487117409706
step: 60, loss: 0.004524665884673595
step: 70, loss: 0.004982091952115297
step: 80, loss: 0.03663714602589607
step: 90, loss: 0.004621097352355719
step: 100, loss: 0.008643526583909988
step: 110, loss: 0.005272530484944582
step: 120, loss: 0.03733668476343155
step: 130, loss: 0.07608113437891006
step: 140, loss: 0.013752521015703678
step: 150, loss: 0.005201738327741623
step: 160, loss: 0.059362299740314484
step: 170, loss: 0.014796700328588486
step: 180, loss: 0.004469507839530706
step: 190, loss: 0.01569812186062336
step: 200, loss: 0.07648266851902008
step: 210, loss: 0.004962523002177477
/usr/local/lib/python3.7/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=0.5872689938398358, f1=0.5987525987525987,
best_f1=0.6205450733752621
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0027959675062447786
step: 10, loss: 0.011509337462484837
step: 20, loss: 0.0333208329975605
step: 30, loss: 0.04504373297095299
step: 40, loss: 0.027171384543180466
step: 50, loss: 0.004705951549112797
step: 60, loss: 0.014954478479921818
step: 70, loss: 0.12411704659461975
step: 80, loss: 0.007474323734641075
step: 90, loss: 0.032051146030426025
step: 100, loss: 0.03170226514339447
step: 110, loss: 0.0011098191607743502
step: 120, loss: 0.023801235482096672
step: 130, loss: 0.015086440369486809

```

```

step: 140, loss: 0.0013513112207874656
step: 150, loss: 0.0006389712798409164
step: 160, loss: 0.014279772527515888
step: 170, loss: 0.0011879106750711799
step: 180, loss: 0.08211299777030945
step: 190, loss: 0.0016763012390583754
step: 200, loss: 0.009314488619565964
step: 210, loss: 0.01569242961704731
/usr/local/lib/python3.7/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=0.5963488843813387, f1=0.6163021868787276,
best_f1=0.6205450733752621
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.021588195115327835
step: 10, loss: 0.0048388345167040825
step: 20, loss: 0.01783786527812481
step: 30, loss: 0.05660362169146538
step: 40, loss: 0.0038432853762060404
step: 50, loss: 0.004351009149104357
step: 60, loss: 0.06054888293147087
step: 70, loss: 0.14339254796504974
step: 80, loss: 0.12448664009571075
step: 90, loss: 0.011294367723166943
step: 100, loss: 0.005223963875323534
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 110, loss: 0.003153391880914569
step: 120, loss: 0.0011111849453300238
step: 130, loss: 0.0032368646934628487
step: 140, loss: 0.002116737887263298
step: 150, loss: 0.009945318102836609
step: 160, loss: 0.01802847534418106
step: 170, loss: 0.09516056627035141
step: 180, loss: 0.03128120303153992
step: 190, loss: 0.002257036743685603
step: 200, loss: 0.0022612162865698338
step: 210, loss: 0.008893809281289577
/usr/local/lib/python3.7/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=0.589430894308943, f1=0.5995975855130784,
best_f1=0.6205450733752621
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.030965182930231094

```

```

step: 10, loss: 0.0011108056642115116
step: 20, loss: 0.003907524980604649
step: 30, loss: 0.001059149857610464
step: 40, loss: 0.0011658815201371908
step: 50, loss: 0.0036210203543305397
step: 60, loss: 0.02451702207326889
step: 70, loss: 0.014927313663065434
step: 80, loss: 0.057624224573373795
step: 90, loss: 0.0025221104733645916
step: 100, loss: 0.0008949787588790059
step: 110, loss: 0.0008089482435025275
step: 120, loss: 0.014427133835852146
step: 130, loss: 0.17281298339366913
step: 140, loss: 0.00611161021515727
step: 150, loss: 0.002893316326662898
step: 160, loss: 0.11141752451658249
step: 170, loss: 0.01980905793607235
step: 180, loss: 0.0032537628430873156
step: 190, loss: 0.0017786030657589436
step: 200, loss: 0.002571376506239176
step: 210, loss: 0.025252999737858772
/usr/local/lib/python3.7/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=0.5891783567134268, f1=0.6104417670682731,
best_f1=0.6205450733752621
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

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

```

```

because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2293it [00:06, 343.97it/s]
load_f1 = 0.6223908918406071
real_f1 = 0.618595825426945
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, 256.98it/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 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

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

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

After processing overrides, optimization options are:

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

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

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

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

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

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

```
step: 0, loss: 0.8710629940032959
```

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

```
step: 10, loss: 0.16800594329833984
```

```
step: 20, loss: 0.15144412219524384
```

```
step: 30, loss: 0.4987943768501282
```

```
step: 40, loss: 0.2526521682739258
```

```
step: 50, loss: 0.30908918380737305
```

```
step: 60, loss: 0.3620380759239197
```

```

step: 70, loss: 0.17383576929569244
step: 80, loss: 0.4927322566509247
step: 90, loss: 0.23529373109340668
step: 100, loss: 0.21641138195991516
step: 110, loss: 0.2383425533771515
step: 120, loss: 0.4167558550834656
step: 130, loss: 0.3537333607673645
step: 140, loss: 0.3057735562324524
step: 150, loss: 0.26579737663269043
step: 160, loss: 0.19649632275104523
step: 170, loss: 0.40037328004837036
step: 180, loss: 0.2310115545988083
step: 190, loss: 0.08577906340360641
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5913461538461539, f1=0.6534653465346535,
best_f1=0.6534653465346535
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3194533586502075
step: 10, loss: 0.019105378538370132
step: 20, loss: 0.07084931433200836
step: 30, loss: 0.20371751487255096
step: 40, loss: 0.46929681301116943
step: 50, loss: 0.33975860476493835
step: 60, loss: 0.08268243074417114
step: 70, loss: 0.15172140300273895
step: 80, loss: 0.14211952686309814
step: 90, loss: 0.11850825697183609
step: 100, loss: 0.2833113670349121
step: 110, loss: 0.09187573194503784
step: 120, loss: 0.40362802147865295
step: 130, loss: 0.15463484823703766
step: 140, loss: 0.1486719399690628
step: 150, loss: 0.029688900336623192
step: 160, loss: 0.021967943757772446
step: 170, loss: 0.09143440425395966
step: 180, loss: 0.24184289574623108
step: 190, loss: 0.17176100611686707
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7048710601719199, f1=0.7192982456140352,
best_f1=0.7192982456140352
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.14376309514045715
step: 10, loss: 0.17826826870441437
step: 20, loss: 0.04981790482997894
step: 30, loss: 0.018709132447838783
step: 40, loss: 0.03989871218800545
step: 50, loss: 0.15948498249053955
step: 60, loss: 0.06790170818567276
step: 70, loss: 0.12011993676424026
step: 80, loss: 0.21796706318855286
step: 90, loss: 0.11446927487850189
step: 100, loss: 0.13260456919670105
step: 110, loss: 0.25052058696746826
step: 120, loss: 0.03366551175713539
step: 130, loss: 0.1275426298379898
step: 140, loss: 0.10034182667732239
step: 150, loss: 0.1499236822128296
step: 160, loss: 0.09748179465532303
step: 170, loss: 0.04970289766788483
step: 180, loss: 0.19930009543895721
step: 190, loss: 0.11123694479465485
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7539267015706806, f1=0.7222222222222222,
best_f1=0.7222222222222222
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014301324263215065
step: 10, loss: 0.11667924374341965
step: 20, loss: 0.005189444869756699
step: 30, loss: 0.057890523225069046
step: 40, loss: 0.0018386347219347954
step: 50, loss: 0.019878128543496132
step: 60, loss: 0.05702534317970276

```



```

step: 70, loss: 0.04345008358359337
step: 80, loss: 0.02525329403579235
step: 90, loss: 0.042356938123703
step: 100, loss: 0.035410843789577484
step: 110, loss: 0.014319914393126965
step: 120, loss: 0.024469919502735138
step: 130, loss: 0.11347516626119614
step: 140, loss: 0.20017938315868378
step: 150, loss: 0.09415119141340256
step: 160, loss: 0.01625113934278488
step: 170, loss: 0.024881020188331604
step: 180, loss: 0.16189253330230713
step: 190, loss: 0.03254012390971184
/usr/local/lib/python3.7/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=0.7486910994764396, f1=0.7388888888888889,
best_f1=0.7222222222222222
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11408045887947083
step: 10, loss: 0.019902285188436508
step: 20, loss: 0.0416107177734375
step: 30, loss: 0.04007617384195328
step: 40, loss: 0.08959970623254776
step: 50, loss: 0.13711318373680115
step: 60, loss: 0.020876821130514145
step: 70, loss: 0.029835013672709465
step: 80, loss: 0.027829866856336594
step: 90, loss: 0.011341406032443047
step: 100, loss: 0.04325539246201515
step: 110, loss: 0.010592259466648102
step: 120, loss: 0.021305963397026062
step: 130, loss: 0.010694680735468864
step: 140, loss: 0.007293000351637602
step: 150, loss: 0.009640702977776527
step: 160, loss: 0.12399724125862122
step: 170, loss: 0.06608381867408752
step: 180, loss: 0.1527264416217804
step: 190, loss: 0.15656836330890656
/usr/local/lib/python3.7/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=0.7493261455525607, f1=0.7344632768361581,
best_f1=0.7222222222222222
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04057863727211952
step: 10, loss: 0.002020695712417364
step: 20, loss: 0.034601807594299316
step: 30, loss: 0.0011649001389741898
step: 40, loss: 0.005260102450847626
step: 50, loss: 0.012650353834033012
step: 60, loss: 0.0005067584570497274
step: 70, loss: 0.0240045003592968
step: 80, loss: 0.04918000474572182
step: 90, loss: 0.00694549223408103
step: 100, loss: 0.018618786707520485
step: 110, loss: 0.03867917135357857
step: 120, loss: 0.010930977761745453
step: 130, loss: 0.007732996717095375
step: 140, loss: 0.0032440146896988153
step: 150, loss: 0.02824297547340393
step: 160, loss: 0.02235395647585392
step: 170, loss: 0.045688778162002563
step: 180, loss: 0.0062300958670675755
step: 190, loss: 0.06375981122255325
/usr/local/lib/python3.7/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=0.75, f1=0.7500000000000001, best_f1=0.7222222222222222
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006933844648301601
step: 10, loss: 0.04266441613435745
step: 20, loss: 0.0006519973394460976
step: 30, loss: 0.028632787987589836
step: 40, loss: 0.14592881500720978
step: 50, loss: 0.029204651713371277
step: 60, loss: 0.06964553147554398
step: 70, loss: 0.0007880173507146537
step: 80, loss: 0.16781756281852722
step: 90, loss: 0.09128567576408386
step: 100, loss: 0.0013193049235269427
step: 110, loss: 0.011524609290063381
step: 120, loss: 0.005917332600802183
step: 130, loss: 0.049391232430934906
step: 140, loss: 0.0026719742454588413
step: 150, loss: 0.03321300819516182
step: 160, loss: 0.0004124938859604299
step: 170, loss: 0.0024349577724933624
step: 180, loss: 0.001638514338992536
step: 190, loss: 0.060898467898368835

```

```

/usr/local/lib/python3.7/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=0.7262872628726287, f1=0.7197802197802198,
best_f1=0.7222222222222222
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009037675918079913
step: 10, loss: 0.015995820984244347
step: 20, loss: 0.0017320243641734123
step: 30, loss: 0.0018152085831388831
step: 40, loss: 0.008869250304996967
step: 50, loss: 0.0011220078449696302
step: 60, loss: 0.002534552477300167
step: 70, loss: 0.004300435073673725
step: 80, loss: 0.015250848606228828
step: 90, loss: 0.028132770210504532
step: 100, loss: 0.037564001977443695
step: 110, loss: 0.003989122342318296
step: 120, loss: 0.06259866058826447
step: 130, loss: 0.004273205064237118
step: 140, loss: 0.001070378115400672
step: 150, loss: 0.0016256263479590416
step: 160, loss: 0.0015402366407215595
step: 170, loss: 0.0010419158497825265
step: 180, loss: 0.0032910609152168036
step: 190, loss: 0.021185332909226418
/usr/local/lib/python3.7/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=0.7481662591687043, f1=0.6945812807881775,
best_f1=0.7222222222222222
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003972738515585661
step: 10, loss: 0.0010587690630927682
step: 20, loss: 0.0038051719311624765
step: 30, loss: 0.005997376050800085
step: 40, loss: 0.0012251463485881686
step: 50, loss: 0.0008425782434642315
step: 60, loss: 0.0005850362940691411
step: 70, loss: 0.001120464876294136
step: 80, loss: 0.012954424135386944
step: 90, loss: 0.0033322402741760015
step: 100, loss: 0.001977470237761736
step: 110, loss: 0.005069221369922161

```

```

step: 120, loss: 0.0035628932528197765
step: 130, loss: 0.000546288734767586
step: 140, loss: 0.00023134253569878638
step: 150, loss: 0.004113460425287485
step: 160, loss: 0.0009682151721790433
step: 170, loss: 0.03346387669444084
step: 180, loss: 0.06441204994916916
step: 190, loss: 0.08539329469203949
/usr/local/lib/python3.7/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=0.7134831460674158, f1=0.7377521613832854,
best_f1=0.7222222222222222
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002303829533047974
step: 10, loss: 0.09151848405599594
step: 20, loss: 0.004372604191303253
step: 30, loss: 0.07610319554805756
step: 40, loss: 0.0026664328761398792
step: 50, loss: 0.0033164448104798794
step: 60, loss: 0.0011004729894921184
step: 70, loss: 0.006981860380619764
step: 80, loss: 0.011543447151780128
step: 90, loss: 0.0012163452338427305
step: 100, loss: 0.002057181904092431
step: 110, loss: 0.01424979418516159
step: 120, loss: 0.012668726034462452
step: 130, loss: 0.0013352162204682827
step: 140, loss: 0.001676853746175766
step: 150, loss: 0.002101736143231392
step: 160, loss: 0.181205153465271
step: 170, loss: 0.01689743436872959
step: 180, loss: 0.0029465199913829565
step: 190, loss: 0.0020994783844798803
/usr/local/lib/python3.7/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=0.7493261455525607, f1=0.7409470752089136,
best_f1=0.7222222222222222
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0023825804237276316
step: 10, loss: 0.0012318699155002832
step: 20, loss: 0.03369037061929703
step: 30, loss: 0.01714910753071308

```

```

step: 40, loss: 0.003036289010196924
step: 50, loss: 0.0005016830400563776
step: 60, loss: 0.001776238321326673
step: 70, loss: 0.0005913199274800718
step: 80, loss: 0.0009426713804714382
step: 90, loss: 0.0023700883612036705
step: 100, loss: 0.0006452550878748298
step: 110, loss: 0.005272555630654097
step: 120, loss: 0.0077692302875220776
step: 130, loss: 0.13079658150672913
step: 140, loss: 0.000711337779648602
step: 150, loss: 0.010954141616821289
step: 160, loss: 0.0008295119041576982
step: 170, loss: 0.0023641656152904034
step: 180, loss: 0.06653555482625961
step: 190, loss: 0.0017533033387735486
/usr/local/lib/python3.7/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=0.7415143603133159, f1=0.7291666666666667,
best_f1=0.7222222222222222
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000632598006632179
step: 10, loss: 0.2037421315908432
step: 20, loss: 0.0038868936244398355
step: 30, loss: 0.004537187051028013
step: 40, loss: 0.04222169145941734
step: 50, loss: 0.004331165924668312
step: 60, loss: 0.0011683576740324497
step: 70, loss: 0.001804353203624487
step: 80, loss: 0.0065732416696846485
step: 90, loss: 0.002259154338389635
step: 100, loss: 0.001464890781790018
step: 110, loss: 0.0006489998195320368
step: 120, loss: 0.002894172677770257
step: 130, loss: 0.0006371706258505583
step: 140, loss: 0.0019507046090438962
step: 150, loss: 0.000992477871477604
step: 160, loss: 0.002195049077272415
step: 170, loss: 0.0022638682276010513
step: 180, loss: 0.0021529209334403276
step: 190, loss: 0.001306951860897243
/usr/local/lib/python3.7/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=0.7513227513227512, f1=0.7238605898123326,

```

```

best_f1=0.7222222222222222
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0020157680846750736
step: 10, loss: 0.000500253401696682
step: 20, loss: 0.018894629552960396
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 30, loss: 0.006345129106193781
step: 40, loss: 0.0009235978941433132
step: 50, loss: 0.018842265009880066
step: 60, loss: 0.000598247570451349
step: 70, loss: 0.0008403699612244964
step: 80, loss: 0.0005425978451967239
step: 90, loss: 0.000739130366127938
step: 100, loss: 0.00121440296061337
step: 110, loss: 0.0005854851915501058
step: 120, loss: 0.0006459032301791012
step: 130, loss: 0.0018175082514062524
step: 140, loss: 0.0013284794986248016
step: 150, loss: 0.03018961101770401
step: 160, loss: 0.004648055415600538
step: 170, loss: 0.00022695593361277133
step: 180, loss: 0.009644201956689358
step: 190, loss: 0.00035318394657224417
/usr/local/lib/python3.7/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=0.7357512953367874, f1=0.7310704960835509,
best_f1=0.7222222222222222
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0019873653072863817
step: 10, loss: 0.0018642593640834093
step: 20, loss: 0.0007305035833269358
step: 30, loss: 0.0022771533112972975
step: 40, loss: 0.0007105585536919534
step: 50, loss: 0.03125208243727684
step: 60, loss: 0.005558728240430355
step: 70, loss: 0.012810434214770794
step: 80, loss: 0.00039583034231327474
step: 90, loss: 0.0011164959287270904
step: 100, loss: 0.14519058167934418
step: 110, loss: 0.0020898908842355013
step: 120, loss: 0.0006690844311378896
step: 130, loss: 0.0016916923923417926
step: 140, loss: 0.0005260429461486638

```

```

step: 150, loss: 0.0003140098706353456
step: 160, loss: 0.00031224661506712437
step: 170, loss: 0.020393967628479004
step: 180, loss: 0.0019444874487817287
step: 190, loss: 0.043679531663656235
/usr/local/lib/python3.7/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=0.7443037974683544, f1=0.7338501291989663,
best_f1=0.7222222222222222
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005449700984172523
step: 10, loss: 0.00126019143499434
step: 20, loss: 0.00036864069988951087
step: 30, loss: 0.00033378825173713267
step: 40, loss: 0.0006198587943799794
step: 50, loss: 0.0013788013020530343
step: 60, loss: 0.001058451016433537
step: 70, loss: 0.0007691921200603247
step: 80, loss: 0.0007046881946735084
step: 90, loss: 0.0007181947003118694
step: 100, loss: 0.0004151703615207225
step: 110, loss: 0.0024360856041312218
step: 120, loss: 0.0012290142476558685
step: 130, loss: 0.0008007683791220188
step: 140, loss: 0.05085853487253189
step: 150, loss: 0.0010450880508869886
step: 160, loss: 0.006142108701169491
step: 170, loss: 0.0351218618452549
step: 180, loss: 0.0015773479826748371
step: 190, loss: 0.002402191748842597
/usr/local/lib/python3.7/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=0.7424242424242425, f1=0.7384615384615385,
best_f1=0.7222222222222222
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

```

```

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True

```

```

master_weights      : True
loss_scale          : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled             : True
opt_level           : 02
cast_model_type     : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights      : True
loss_scale          : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2049it [00:09, 222.18it/s]
load_f1 = 0.6038647342995169
real_f1 = 0.5920745920745921
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]

```

1.2 DA TEXTUAL

1.2.1 Abt-Buy - Running the matcher

```

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

```



```

--da entry_swap \
--da span_shuffle \
--summarize

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

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

```

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

Defaults for this optimization level are:

```

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

```

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

After processing overrides, optimization options are:

```

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

```

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

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

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

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

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

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

step: 0, loss: 0.85451740026474

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

step: 10, loss: 0.23021988570690155

step: 20, loss: 0.15086327493190765

step: 30, loss: 0.23604591190814972

step: 40, loss: 0.31853461265563965

step: 50, loss: 0.3843930959701538

step: 60, loss: 0.4417169690132141

step: 70, loss: 0.3073020875453949

step: 80, loss: 0.24551329016685486

step: 90, loss: 0.39365091919898987

step: 100, loss: 0.22860506176948547

step: 110, loss: 0.17610566318035126

step: 120, loss: 0.5523548722267151

step: 130, loss: 0.4073079824447632

step: 140, loss: 0.4248028099536896

step: 150, loss: 0.06423655152320862

step: 160, loss: 0.18669702112674713

step: 170, loss: 0.11764011532068253

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

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

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

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=0.6979166666666666, f1=0.7010309278350515,

best_f1=0.7010309278350515

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

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

step: 0, loss: 0.19306057691574097

step: 10, loss: 0.12137340009212494

step: 20, loss: 0.22770848870277405

step: 30, loss: 0.1863856315612793

step: 40, loss: 0.16291698813438416

step: 50, loss: 0.14910218119621277

step: 60, loss: 0.09581246972084045

step: 70, loss: 0.1804276406764984

step: 80, loss: 0.1340898722410202

step: 90, loss: 0.2913558781147003

step: 100, loss: 0.23192015290260315

```

step: 110, loss: 0.14255943894386292
step: 120, loss: 0.052628714591264725
step: 130, loss: 0.048974718898534775
step: 140, loss: 0.10822061449289322
step: 150, loss: 0.1106770858168602
step: 160, loss: 0.0924074575304985
step: 170, loss: 0.04897609353065491
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7291666666666666, f1=0.7213930348258706,
best_f1=0.7213930348258706
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04696735739707947
step: 10, loss: 0.03131505846977234
step: 20, loss: 0.018038470298051834
step: 30, loss: 0.11931087076663971
step: 40, loss: 0.03299550339579582
step: 50, loss: 0.12593427300453186
step: 60, loss: 0.10925605148077011
step: 70, loss: 0.07698006927967072
step: 80, loss: 0.12339560687541962
step: 90, loss: 0.06922166049480438
step: 100, loss: 0.03725428879261017
step: 110, loss: 0.08310443162918091
step: 120, loss: 0.010603472590446472
step: 130, loss: 0.07592957466840744
step: 140, loss: 0.004047134891152382
step: 150, loss: 0.19804102182388306
step: 160, loss: 0.02737879566848278
step: 170, loss: 0.1199922263622284
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.8041237113402061, f1=0.7758186397984886,
best_f1=0.7758186397984886
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0408005528151989
step: 10, loss: 0.27568739652633667
step: 20, loss: 0.03790696710348129
step: 30, loss: 0.030353853479027748
step: 40, loss: 0.10415805131196976
step: 50, loss: 0.030807124450802803
step: 60, loss: 0.10617301613092422
step: 70, loss: 0.09085168689489365
step: 80, loss: 0.03361966833472252
step: 90, loss: 0.029735011979937553
step: 100, loss: 0.02654142491519451
step: 110, loss: 0.04252532124519348
step: 120, loss: 0.1391245424747467
step: 130, loss: 0.05403359606862068
step: 140, loss: 0.027670051902532578
step: 150, loss: 0.004687822423875332
step: 160, loss: 0.057168517261743546
step: 170, loss: 0.006761020980775356
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8155844155844156, f1=0.8089330024813897,
best_f1=0.8089330024813897
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012547746300697327
step: 10, loss: 0.058019496500492096
step: 20, loss: 0.039771080017089844
step: 30, loss: 0.05992919206619263
step: 40, loss: 0.029280751943588257
step: 50, loss: 0.005397832952439785
step: 60, loss: 0.0006125308573246002
step: 70, loss: 0.033729128539562225
step: 80, loss: 0.0026247152127325535
step: 90, loss: 0.05297999083995819
step: 100, loss: 0.048224300146102905
step: 110, loss: 0.08711379766464233
step: 120, loss: 0.02683999575674534
step: 130, loss: 0.00144485617056489
step: 140, loss: 0.039442528039216995
step: 150, loss: 0.04452136158943176
step: 160, loss: 0.08291604369878769

```

```

step: 170, loss: 0.1308431774377823
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8210526315789473, f1=0.815, best_f1=0.815
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.022790249437093735
step: 10, loss: 0.10786548256874084
step: 20, loss: 0.0762673020362854
step: 30, loss: 0.09921828657388687
step: 40, loss: 0.0021342108957469463
step: 50, loss: 0.03477305918931961
step: 60, loss: 0.040424786508083344
step: 70, loss: 0.01305882353335619
step: 80, loss: 0.051353421062231064
step: 90, loss: 0.022319791838526726
step: 100, loss: 0.015094488859176636
step: 110, loss: 0.101203553378582
step: 120, loss: 0.01919180527329445
step: 130, loss: 0.1667805314064026
step: 140, loss: 0.004161223769187927
step: 150, loss: 0.16212545335292816
step: 160, loss: 0.035401709377765656
step: 170, loss: 0.0007586986757814884
/usr/local/lib/python3.7/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=0.8188976377952756, f1=0.7969543147208121, best_f1=0.815
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001154051162302494
step: 10, loss: 0.0013924404047429562
step: 20, loss: 0.01725267432630062
step: 30, loss: 0.0021479015704244375
step: 40, loss: 0.007749613840132952
step: 50, loss: 0.0018111328827217221
step: 60, loss: 0.06969068199396133
step: 70, loss: 0.047593727707862854
step: 80, loss: 0.006013545673340559
step: 90, loss: 0.02551833726465702
step: 100, loss: 0.011267648078501225

```

```

step: 110, loss: 0.0012015114771202207
step: 120, loss: 0.3081480860710144
step: 130, loss: 0.04211508110165596
step: 140, loss: 0.041937075555324554
step: 150, loss: 0.07509899884462357
step: 160, loss: 0.0198604017496109
step: 170, loss: 0.04591864347457886
/usr/local/lib/python3.7/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=0.8073878627968338, f1=0.7909319899244334, best_f1=0.815
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.034472472965717316
step: 10, loss: 0.013571714051067829
step: 20, loss: 0.0012819991679862142
step: 30, loss: 0.006187898572534323
step: 40, loss: 0.0016185167478397489
step: 50, loss: 0.001972862519323826
step: 60, loss: 0.0023622713051736355
step: 70, loss: 0.019375033676624298
step: 80, loss: 0.0003753257915377617
step: 90, loss: 0.006196115631610155
step: 100, loss: 0.0048536635003983974
step: 110, loss: 0.009642787277698517
step: 120, loss: 0.0018514868570491672
step: 130, loss: 0.0005711138364858925
step: 140, loss: 0.039705220609903336
step: 150, loss: 0.040977321565151215
step: 160, loss: 0.02014811895787716
step: 170, loss: 0.00540338596329093
/usr/local/lib/python3.7/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=0.8193384223918575, f1=0.7970660146699265, best_f1=0.815
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009210366755723953
step: 10, loss: 0.002390759764239192
step: 20, loss: 0.006535208318382502
step: 30, loss: 0.032884836196899414
step: 40, loss: 0.021237043663859367
step: 50, loss: 0.003185876877978444
step: 60, loss: 0.00044587955926544964
step: 70, loss: 0.002067109802737832
step: 80, loss: 0.03449880704283714

```

```

step: 90, loss: 0.00881605502218008
step: 100, loss: 0.07627008110284805
step: 110, loss: 0.00010603380360407755
step: 120, loss: 0.05019068345427513
step: 130, loss: 0.0015907760243862867
step: 140, loss: 0.0008196506532840431
step: 150, loss: 0.009207522496581078
step: 160, loss: 0.01715158112347126
step: 170, loss: 0.0427774153649807
/usr/local/lib/python3.7/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=0.8052631578947368, f1=0.7857142857142857, best_f1=0.815
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005801809020340443
step: 10, loss: 0.04231301322579384
step: 20, loss: 0.00017203562310896814
step: 30, loss: 0.000142794189741835
step: 40, loss: 0.03931897506117821
step: 50, loss: 0.026459895074367523
step: 60, loss: 0.005705576390028
step: 70, loss: 0.0031703694257885218
step: 80, loss: 0.03179777041077614
step: 90, loss: 0.11585675179958344
step: 100, loss: 0.05749146267771721
step: 110, loss: 0.011032958514988422
step: 120, loss: 0.013240362517535686
step: 130, loss: 0.0024044064339250326
step: 140, loss: 0.11073829233646393
step: 150, loss: 0.0016582024982199073
step: 160, loss: 0.0009590999106876552
step: 170, loss: 0.004125380888581276
/usr/local/lib/python3.7/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=0.8108108108108109, f1=0.7989556135770236, best_f1=0.815
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009272953029721975
step: 10, loss: 0.0034548346884548664
step: 20, loss: 0.06590621173381805
step: 30, loss: 0.003870979882776737
step: 40, loss: 0.007920365780591965
step: 50, loss: 0.04309302568435669
step: 60, loss: 0.007576403208076954

```

```

step: 70, loss: 0.0007511815056204796
step: 80, loss: 0.00037537762545980513
step: 90, loss: 0.005777830258011818
step: 100, loss: 0.0007163650006987154
step: 110, loss: 0.0028002236504107714
step: 120, loss: 0.010692263022065163
step: 130, loss: 0.004429645370692015
step: 140, loss: 0.04362424090504646
step: 150, loss: 0.01074772048741579
step: 160, loss: 0.0008672783151268959
step: 170, loss: 0.0932152196764946
/usr/local/lib/python3.7/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=0.8128342245989305, f1=0.7918781725888325, best_f1=0.815
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0038640405982732773
step: 10, loss: 0.00960404984652996
step: 20, loss: 0.024123217910528183
step: 30, loss: 0.030502311885356903
step: 40, loss: 0.00019507165416143835
step: 50, loss: 0.0002704166399780661
step: 60, loss: 0.0012608885299414396
step: 70, loss: 0.0053762891329824924
step: 80, loss: 0.0059226988814771175
step: 90, loss: 0.00015370361506938934
step: 100, loss: 0.001476371195167303
step: 110, loss: 6.485976336989552e-05
step: 120, loss: 0.0020139238331466913
step: 130, loss: 0.0012001401046290994
step: 140, loss: 0.022076083347201347
step: 150, loss: 0.028309473767876625
step: 160, loss: 0.0010508436243981123
step: 170, loss: 0.0033488026820123196
/usr/local/lib/python3.7/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=0.8083989501312336, f1=0.797979797979798, best_f1=0.815
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00038861323264427483
step: 10, loss: 0.050359904766082764
step: 20, loss: 0.0008326454553753138
step: 30, loss: 0.0013091262662783265
step: 40, loss: 0.00010892251884797588

```



```

step: 50, loss: 0.02492223121225834
step: 60, loss: 0.020243503153324127
step: 70, loss: 0.0006060847663320601
step: 80, loss: 0.0001655383821344003
step: 90, loss: 0.00018245153478346765
step: 100, loss: 6.95664421073161e-05
step: 110, loss: 0.022020163014531136
step: 120, loss: 0.0003772558702621609
step: 130, loss: 0.007681798655539751
step: 140, loss: 0.0012243095552548766
step: 150, loss: 0.0001993402693187818
step: 160, loss: 0.0016069895355030894
step: 170, loss: 0.0005853091715835035
/usr/local/lib/python3.7/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=0.8054054054054054, f1=0.7897435897435897, best_f1=0.815
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00022518336481880397
step: 10, loss: 7.382001058431342e-05
step: 20, loss: 0.008724833838641644
step: 30, loss: 0.0029959646053612232
step: 40, loss: 0.0015334865311160684
step: 50, loss: 0.0008990225032903254
step: 60, loss: 9.410289931111038e-05
step: 70, loss: 0.0160381980240345
step: 80, loss: 0.000252304773312062
step: 90, loss: 0.00131186842918396
step: 100, loss: 0.0002166948252124712
step: 110, loss: 0.0013365132035687566
step: 120, loss: 0.0001994376943912357
step: 130, loss: 0.00021274908795021474
step: 140, loss: 0.005674657411873341
step: 150, loss: 0.0006414036033675075
step: 160, loss: 0.03394407406449318
step: 170, loss: 0.00013933185255154967
/usr/local/lib/python3.7/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=0.8073878627968338, f1=0.7890818858560793, best_f1=0.815
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0022526420652866364
step: 10, loss: 0.0015414864756166935
step: 20, loss: 0.015234916470944881

```

```

step: 30, loss: 0.006285760551691055
step: 40, loss: 0.00032097107032313943
step: 50, loss: 0.0005515097291208804
step: 60, loss: 0.0001519021752756089
step: 70, loss: 0.0007623771671205759
step: 80, loss: 0.0003930207167286426
step: 90, loss: 8.656705176690593e-05
step: 100, loss: 0.0006985988584347069
step: 110, loss: 0.00010487234976608306
step: 120, loss: 0.0035484882537275553
step: 130, loss: 0.0010842025512829423
step: 140, loss: 0.0021297442726790905
step: 150, loss: 0.06331441551446915
step: 160, loss: 9.098913142224774e-05
step: 170, loss: 0.0022301520220935345
/usr/local/lib/python3.7/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=0.8073878627968338, f1=0.7890818858560793, best_f1=0.815
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

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

```

```

    warnings.warn("An input tensor was not cuda.")
1916it [00:05, 329.72it/s]
load_f1 = 0.6204081632653061
real_f1 = 0.5698529411764707
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, 251.38it/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, 547kB/s]
Downloading: 100% 232k/232k [00:00<00:00, 1.22MB/s]

```

Downloading: 100% 268M/268M [00:05<00:00, 48.6MB/s]
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

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

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

After processing overrides, optimization options are:

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

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

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

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

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

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

step: 0, loss: 0.821397066116333
step: 10, loss: 0.4536892771720886
step: 20, loss: 0.5559422373771667
step: 30, loss: 0.3426201045513153
step: 40, loss: 0.14525994658470154
step: 50, loss: 0.05682886764407158
step: 60, loss: 0.18349668383598328
step: 70, loss: 0.11398960649967194
step: 80, loss: 0.18219804763793945
step: 90, loss: 0.0750570148229599
step: 100, loss: 0.08570762723684311

```

step: 110, loss: 0.09258750081062317
step: 120, loss: 0.024425897747278214
step: 130, loss: 0.009818150661885738
step: 140, loss: 0.02277599461376667
step: 150, loss: 0.14789189398288727
step: 160, loss: 0.25485360622406006
step: 170, loss: 0.013921873643994331
step: 180, loss: 0.011134189553558826
step: 190, loss: 0.02210364118218422
step: 200, loss: 0.006882947403937578
step: 210, loss: 0.021048234775662422
step: 220, loss: 0.007217614911496639
step: 230, loss: 0.048737719655036926
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9617977528089887, f1=0.960272417707151,
best_f1=0.960272417707151
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03405741974711418
step: 10, loss: 0.008691342547535896
step: 20, loss: 0.006112925708293915
step: 30, loss: 0.0039899940602481365
step: 40, loss: 0.016003187745809555
step: 50, loss: 0.028953690081834793
step: 60, loss: 0.00502352137118578
step: 70, loss: 0.010550437495112419
step: 80, loss: 0.0036862122360616922
step: 90, loss: 0.01544682215899229
step: 100, loss: 0.1316918432712555
step: 110, loss: 0.06266989558935165
step: 120, loss: 0.010378042235970497
step: 130, loss: 0.001754004042595625
step: 140, loss: 0.20566806197166443
step: 150, loss: 0.010397669859230518
step: 160, loss: 0.008714203722774982
step: 170, loss: 0.06436160206794739
step: 180, loss: 0.0038445403333753347
step: 190, loss: 0.0871000811457634
step: 200, loss: 0.02364843338727951
step: 210, loss: 0.10425562411546707
step: 220, loss: 0.00412218039855361

```

```

step: 230, loss: 0.014010405167937279
/usr/local/lib/python3.7/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=0.9568106312292359, f1=0.9608062709966406,
best_f1=0.960272417707151
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.14585532248020172
step: 10, loss: 0.01570853590965271
step: 20, loss: 0.008411193266510963
step: 30, loss: 0.030057763680815697
step: 40, loss: 0.0661320611834526
step: 50, loss: 0.004491456784307957
step: 60, loss: 0.011042232625186443
step: 70, loss: 0.0039956276305019855
step: 80, loss: 0.01840619556605816
step: 90, loss: 0.018630390986800194
step: 100, loss: 0.0016056440072134137
step: 110, loss: 0.061095282435417175
step: 120, loss: 0.029292721301317215
step: 130, loss: 0.001696280320174992
step: 140, loss: 0.01583457738161087
step: 150, loss: 0.007060921285301447
step: 160, loss: 0.0014802698278799653
step: 170, loss: 0.003313458291813731
step: 180, loss: 0.018772536888718605
step: 190, loss: 0.005595518741756678
step: 200, loss: 0.018191276118159294
step: 210, loss: 0.007201375439763069
step: 220, loss: 0.004251627717167139
step: 230, loss: 0.10260595381259918
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.963718820861678,
best_f1=0.963718820861678
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010015126317739487
step: 10, loss: 0.003358451882377267
step: 20, loss: 0.0012146384688094258

```

```

step: 30, loss: 0.000751134823076427
step: 40, loss: 0.03918331116437912
step: 50, loss: 0.002483104355633259
step: 60, loss: 0.012215427123010159
step: 70, loss: 0.21992594003677368
step: 80, loss: 0.05426914617419243
step: 90, loss: 0.0023023809771984816
step: 100, loss: 0.0031941328197717667
step: 110, loss: 0.008840758353471756
step: 120, loss: 0.003985400777310133
step: 130, loss: 0.0308679286390543
step: 140, loss: 0.0013530324213206768
step: 150, loss: 0.01217762753367424
step: 160, loss: 0.0016080399509519339
step: 170, loss: 0.0058933161199092865
step: 180, loss: 0.11514127254486084
step: 190, loss: 0.010695826262235641
step: 200, loss: 0.02094029262661934
step: 210, loss: 0.002104908460751176
step: 220, loss: 0.0017949981847777963
step: 230, loss: 0.0006930715753696859
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9753363228699552, f1=0.9740112994350283,
best_f1=0.9740112994350283
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0031074965372681618
step: 10, loss: 0.0019519588677212596
step: 20, loss: 0.0011437685461714864
step: 30, loss: 0.0011545912129804492
step: 40, loss: 0.0023510006722062826
step: 50, loss: 0.00043936484144069254
step: 60, loss: 0.000557230319827795
step: 70, loss: 0.0005790952709503472
step: 80, loss: 0.0008851656457409263
step: 90, loss: 0.002471935236826539
step: 100, loss: 0.006215553265064955
step: 110, loss: 0.0005717876483686268
step: 120, loss: 0.043946120887994766
step: 130, loss: 0.0038357514422386885
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0

```

```

step: 140, loss: 0.0007315102848224342
step: 150, loss: 0.0009424650925211608
step: 160, loss: 0.037996403872966766
step: 170, loss: 0.0033682335633784533
step: 180, loss: 0.0014268800150603056
step: 190, loss: 0.0006524728960357606
step: 200, loss: 0.011440007947385311
step: 210, loss: 0.0005441382527351379
step: 220, loss: 0.0007922840886749327
step: 230, loss: 0.0050205145962536335
/usr/local/lib/python3.7/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=0.9683257918552037, f1=0.9703872437357631,
best_f1=0.9740112994350283
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007908548228442669
step: 10, loss: 0.04982811585068703
step: 20, loss: 0.0007084172684699297
step: 30, loss: 0.00022554348106496036
step: 40, loss: 0.0023625153116881847
step: 50, loss: 0.06693913042545319
step: 60, loss: 0.0011494411155581474
step: 70, loss: 0.000710606575012207
step: 80, loss: 0.0010770132066681981
step: 90, loss: 0.008699720725417137
step: 100, loss: 0.0008139819256030023
step: 110, loss: 0.022856151685118675
step: 120, loss: 0.0011849418515339494
step: 130, loss: 0.0004629720642697066
step: 140, loss: 0.060673899948596954
step: 150, loss: 0.0010272909421473742
step: 160, loss: 0.010313771665096283
step: 170, loss: 0.0007529532304033637
step: 180, loss: 0.0010458002798259258
step: 190, loss: 0.09673715382814407
step: 200, loss: 0.005988029297441244
step: 210, loss: 0.005173408892005682
step: 220, loss: 0.014266782440245152
step: 230, loss: 0.004743678029626608
/usr/local/lib/python3.7/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=0.9710467706013363, f1=0.9719416386083053,
best_f1=0.9740112994350283
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```



```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11291896551847458
step: 10, loss: 0.002233409322798252
step: 20, loss: 0.0009818641701713204
step: 30, loss: 0.001213033334352076
step: 40, loss: 0.001846611499786377
step: 50, loss: 0.00030998000875115395
step: 60, loss: 0.007078276015818119
step: 70, loss: 0.0020289658568799496
step: 80, loss: 0.00041080446681007743
step: 90, loss: 0.0027567392680794
step: 100, loss: 0.00141475023701787
step: 110, loss: 0.0040305303409695625
step: 120, loss: 0.0024246128741651773
step: 130, loss: 0.005982753820717335
step: 140, loss: 0.00015536861610598862
step: 150, loss: 0.011034676805138588
step: 160, loss: 0.0004930945578962564
step: 170, loss: 0.00026331646949984133
step: 180, loss: 0.00022371308295987546
step: 190, loss: 0.003955169580876827
step: 200, loss: 0.0006988988607190549
step: 210, loss: 0.0004544803232420236
step: 220, loss: 0.0002687110973056406
step: 230, loss: 0.014803188852965832
/usr/local/lib/python3.7/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=0.9730337078651685, f1=0.9718785151856018,
best_f1=0.9740112994350283
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.015387208200991154
step: 10, loss: 0.016082637012004852
step: 20, loss: 0.00025898858439177275
step: 30, loss: 0.000572427932638675
step: 40, loss: 0.0006334945792332292
step: 50, loss: 0.09531362354755402
step: 60, loss: 0.00030197721207514405
step: 70, loss: 0.00029365296359173954
step: 80, loss: 0.002900599967688322
step: 90, loss: 0.00028281519189476967
step: 100, loss: 0.0004747846396639943
step: 110, loss: 0.002140807220712304
step: 120, loss: 8.220254676416516e-05
step: 130, loss: 0.007727850694209337

```

```

step: 140, loss: 0.0001733476237859577
step: 150, loss: 0.00010621689580148086
step: 160, loss: 0.0006241776281967759
step: 170, loss: 0.0003223121166229248
step: 180, loss: 0.0022004873026162386
step: 190, loss: 0.010661964304745197
step: 200, loss: 0.0030809850431978703
step: 210, loss: 0.0008292844868265092
step: 220, loss: 0.00020745811343658715
step: 230, loss: 0.08255413174629211
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.976271186440678, f1=0.963718820861678,
best_f1=0.963718820861678
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005964712472632527
step: 10, loss: 0.0025202722754329443
step: 20, loss: 0.004814933054149151
step: 30, loss: 0.00249082432128489
step: 40, loss: 0.00012837638496421278
step: 50, loss: 0.000149432904436253
step: 60, loss: 0.00033347573480568826
step: 70, loss: 0.0005535150412470102
step: 80, loss: 0.04177550598978996
step: 90, loss: 0.0009582744678482413
step: 100, loss: 0.0004943316453136504
step: 110, loss: 0.0033578851725906134
step: 120, loss: 0.036349713802337646
step: 130, loss: 0.00012868837802670896
step: 140, loss: 0.02672968991100788
step: 150, loss: 0.00011375731264706701
step: 160, loss: 0.0001920007634907961
step: 170, loss: 0.00010337740241084248
step: 180, loss: 0.0002419582160655409
step: 190, loss: 7.856877346057445e-05
step: 200, loss: 0.00017399837088305503
step: 210, loss: 0.00011088239261880517
step: 220, loss: 0.028247766196727753
step: 230, loss: 0.003943862393498421
/usr/local/lib/python3.7/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=0.9740112994350283, f1=0.9704545454545453,
best_f1=0.963718820861678
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00016629850142635405
step: 10, loss: 0.001072760671377182
step: 20, loss: 9.049275104189292e-05
step: 30, loss: 8.80362349562347e-05
step: 40, loss: 7.224169530672953e-05
step: 50, loss: 0.000704327889252454
step: 60, loss: 0.08457424491643906
step: 70, loss: 0.0005341109936125576
step: 80, loss: 0.00023526549921371043
step: 90, loss: 0.00012620989582501352
step: 100, loss: 9.13605181267485e-05
step: 110, loss: 0.00015284348046407104
step: 120, loss: 0.023249808698892593
step: 130, loss: 0.0003437263367231935
step: 140, loss: 0.06648097187280655
step: 150, loss: 0.0003577622410375625
step: 160, loss: 0.0019721314311027527
step: 170, loss: 0.00027300312649458647
step: 180, loss: 0.00028903488419018686
step: 190, loss: 0.0003418927954044193
step: 200, loss: 8.725041698198766e-05
step: 210, loss: 0.0003365033771842718
step: 220, loss: 0.005815885029733181
step: 230, loss: 0.021729858592152596
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9773755656108598, f1=0.9659863945578231,
best_f1=0.9659863945578231
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.284792031394318e-05
step: 10, loss: 0.0004789596132468432
step: 20, loss: 4.426489977049641e-05
step: 30, loss: 0.00022906981757842004
step: 40, loss: 0.015849346294999123
step: 50, loss: 0.000868248229380697

```

```

step: 60, loss: 0.00024177321756724268
step: 70, loss: 0.00010815149289555848
step: 80, loss: 0.015151221305131912
step: 90, loss: 0.00012370356125757098
step: 100, loss: 0.00010406727233203128
step: 110, loss: 0.0002509920741431415
step: 120, loss: 7.19667732482776e-05
step: 130, loss: 4.115116098546423e-05
step: 140, loss: 9.422446601092815e-05
step: 150, loss: 5.5606313253520057e-05
step: 160, loss: 7.948310667416081e-05
step: 170, loss: 0.007738485466688871
step: 180, loss: 9.571816917741671e-05
step: 190, loss: 0.0002778155030682683
step: 200, loss: 0.00037131947465240955
step: 210, loss: 6.468495848821476e-05
step: 220, loss: 8.729607361601666e-05
step: 230, loss: 0.007320572156459093
/usr/local/lib/python3.7/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=0.9732739420935412, f1=0.968609865470852,
best_f1=0.9659863945578231
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006767398444935679
step: 10, loss: 0.0006982986233197153
step: 20, loss: 5.0798076699720696e-05
step: 30, loss: 0.008509315550327301
step: 40, loss: 8.776308095548302e-05
step: 50, loss: 0.0034196695778518915
step: 60, loss: 0.0021708174608647823
step: 70, loss: 0.00011614149843808264
step: 80, loss: 6.218224734766409e-05
step: 90, loss: 7.774307596264407e-05
step: 100, loss: 5.7073848438449204e-05
step: 110, loss: 0.00022193125914782286
step: 120, loss: 0.0005968693876639009
step: 130, loss: 0.0001347100333077833
step: 140, loss: 0.00017015599587466568
step: 150, loss: 5.2172017603879794e-05
step: 160, loss: 0.05071074515581131
step: 170, loss: 7.459439802914858e-05
step: 180, loss: 0.00017813385056797415
step: 190, loss: 8.144199819071218e-05
step: 200, loss: 0.012999288737773895
step: 210, loss: 7.003066275501624e-05

```

```

step: 220, loss: 0.0037070519756525755
step: 230, loss: 5.624693585559726e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 12: dev_f1=0.9774266365688488, f1=0.9683257918552037,
best_f1=0.9683257918552037
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.859875717666e-05
step: 10, loss: 6.916665006428957e-05
step: 20, loss: 0.000176224930328317
step: 30, loss: 0.003352060914039612
step: 40, loss: 0.00018991372780874372
step: 50, loss: 6.54322502668947e-05
step: 60, loss: 5.325058737071231e-05
step: 70, loss: 5.909458559472114e-05
step: 80, loss: 3.059843947994523e-05
step: 90, loss: 3.253926115576178e-05
step: 100, loss: 5.215115743339993e-05
step: 110, loss: 4.7007033572299406e-05
step: 120, loss: 8.048710878938437e-05
step: 130, loss: 6.627121183555573e-05
step: 140, loss: 0.03605102375149727
step: 150, loss: 0.0045061842538416386
step: 160, loss: 6.426362233469263e-05
step: 170, loss: 7.114804611774161e-05
step: 180, loss: 0.00017497886437922716
step: 190, loss: 6.0100363043602556e-05
step: 200, loss: 7.578789518447593e-05
step: 210, loss: 5.928000609856099e-05
step: 220, loss: 3.6848730815108865e-05
step: 230, loss: 3.18127749778796e-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=0.9743589743589743, f1=0.9708520179372198,
best_f1=0.9683257918552037
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.423674494726583e-05
step: 10, loss: 2.948456494777929e-05

```

```

step: 20, loss: 0.0001609826140338555
step: 30, loss: 0.00027406931621953845
step: 40, loss: 5.8597655879566446e-05
step: 50, loss: 6.659844075329602e-05
step: 60, loss: 2.6609042834024876e-05
step: 70, loss: 5.19896020705346e-05
step: 80, loss: 0.00013053734437562525
step: 90, loss: 6.280629168031737e-05
step: 100, loss: 0.010200762189924717
step: 110, loss: 5.97979269514326e-05
step: 120, loss: 4.430492117535323e-05
step: 130, loss: 8.083565626293421e-05
step: 140, loss: 0.0014668996445834637
step: 150, loss: 6.96196366334334e-05
step: 160, loss: 3.373099389136769e-05
step: 170, loss: 7.306494080694392e-05
step: 180, loss: 7.564694533357397e-05
step: 190, loss: 0.002336986595764756
step: 200, loss: 4.179949974059127e-05
step: 210, loss: 0.0003159190819133073
step: 220, loss: 0.0002924844447989017
step: 230, loss: 2.3770626285113394e-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=0.976271186440678, f1=0.9682539682539683,
best_f1=0.9683257918552037
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.179724293062463e-05
step: 10, loss: 7.679632108192891e-05
step: 20, loss: 0.00014614891551900655
step: 30, loss: 0.00011685498611768708
step: 40, loss: 5.3379961173050106e-05
step: 50, loss: 7.630138861713931e-05
step: 60, loss: 3.363061114214361e-05
step: 70, loss: 0.00020769477123394608
step: 80, loss: 3.6465378798311576e-05
step: 90, loss: 3.281039244029671e-05
step: 100, loss: 5.3260799177223817e-05
step: 110, loss: 5.196053461986594e-05
step: 120, loss: 8.86850684764795e-05
step: 130, loss: 6.167680840007961e-05
step: 140, loss: 0.0003181250940542668
step: 150, loss: 0.00010069309792015702
step: 160, loss: 0.0001811045513022691
step: 170, loss: 4.636043740902096e-05

```

```

step: 180, loss: 4.255912426742725e-05
step: 190, loss: 0.0019959884230047464
step: 200, loss: 3.796643431996927e-05
step: 210, loss: 0.0006556982407346368
step: 220, loss: 0.0010473916772753
step: 230, loss: 6.98986477800645e-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=0.9762174405436014, f1=0.9681818181818181,
best_f1=0.9683257918552037
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2473it [00:07, 314.10it/s]
load_f1 = 0.976271186440678
real_f1 = 0.9751693002257337
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, 355.68it/s]

```

1.3.2 DBLP-GoogleScholar - Running the matcher

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

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

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

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

Defaults for this optimization level are:

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

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

After processing overrides, optimization options are:

```
enabled          : True
```



```

opt_level                : 02
cast_model_type          : torch.float16
patch_torch_functions    : False
keep_batchnorm_fp32      : True
master_weights           : True
loss_scale               : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8000061511993408
step: 10, loss: 0.4193602204322815
step: 20, loss: 0.5060062408447266
step: 30, loss: 0.41839492321014404
step: 40, loss: 0.3191046714782715
step: 50, loss: 0.18436148762702942
step: 60, loss: 0.15770366787910461
step: 70, loss: 0.1364724487066269
step: 80, loss: 0.13346101343631744
step: 90, loss: 0.16894683241844177
step: 100, loss: 0.3221724033355713
step: 110, loss: 0.09482378512620926
step: 120, loss: 0.031715430319309235
step: 130, loss: 0.04621207341551781
step: 140, loss: 0.10150561481714249
step: 150, loss: 0.04195712134242058
step: 160, loss: 0.16062021255493164
step: 170, loss: 0.08925167471170425
step: 180, loss: 0.12400756031274796
step: 190, loss: 0.05769435688853264
step: 200, loss: 0.13963846862316132
step: 210, loss: 0.08528351038694382
step: 220, loss: 0.11236830800771713
step: 230, loss: 0.14201045036315918
step: 240, loss: 0.053174152970314026
step: 250, loss: 0.04273369163274765

```

```

step: 260, loss: 0.03415720909833908
step: 270, loss: 0.03585418686270714
step: 280, loss: 0.11525171995162964
step: 290, loss: 0.05263613536953926
step: 300, loss: 0.14001955091953278
step: 310, loss: 0.14348316192626953
step: 320, loss: 0.05577575042843819
step: 330, loss: 0.1505858451128006
step: 340, loss: 0.1777513027191162
step: 350, loss: 0.12048815935850143
step: 360, loss: 0.07548493146896362
step: 370, loss: 0.1780814677476883
step: 380, loss: 0.20113477110862732
step: 390, loss: 0.12157560139894485
step: 400, loss: 0.05018192529678345
step: 410, loss: 0.03375035896897316
step: 420, loss: 0.007337137125432491
step: 430, loss: 0.0808197483420372
step: 440, loss: 0.07248810678720474
step: 450, loss: 0.05583111569285393
step: 460, loss: 0.19896365702152252
step: 470, loss: 0.29033780097961426
step: 480, loss: 0.19829881191253662
step: 490, loss: 0.05787067115306854
step: 500, loss: 0.0346657820045948
step: 510, loss: 0.033010900020599365
step: 520, loss: 0.11239591240882874
step: 530, loss: 0.10143748670816422
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.921451538814883, f1=0.9195933456561923,
best_f1=0.9195933456561923
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1504153609275818
step: 10, loss: 0.21433253586292267
step: 20, loss: 0.16201643645763397
step: 30, loss: 0.09309521317481995
step: 40, loss: 0.02158311940729618
step: 50, loss: 0.036039985716342926
step: 60, loss: 0.18748772144317627
step: 70, loss: 0.1245678961277008

```

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

step: 80, loss: 0.06114480644464493
step: 90, loss: 0.0017730018589645624
step: 100, loss: 0.20604397356510162
step: 110, loss: 0.03709324821829796
step: 120, loss: 0.06675592809915543
step: 130, loss: 0.02237275056540966
step: 140, loss: 0.02762768231332302
step: 150, loss: 0.13167008757591248
step: 160, loss: 0.06595714390277863
step: 170, loss: 0.20640401542186737
step: 180, loss: 0.06833847612142563
step: 190, loss: 0.0291915126144886
step: 200, loss: 0.03562544286251068
step: 210, loss: 0.04380076006054878
step: 220, loss: 0.20297855138778687
step: 230, loss: 0.08999712765216827
step: 240, loss: 0.13168707489967346
step: 250, loss: 0.03869815170764923
step: 260, loss: 0.017198581248521805
step: 270, loss: 0.07601851224899292
step: 280, loss: 0.19856253266334534
step: 290, loss: 0.059574853628873825
step: 300, loss: 0.008892097510397434
step: 310, loss: 0.06548071652650833
step: 320, loss: 0.27299362421035767
step: 330, loss: 0.02165369503200054
step: 340, loss: 0.012508586049079895
step: 350, loss: 0.03433249518275261
step: 360, loss: 0.05052027478814125
step: 370, loss: 0.008521503768861294
step: 380, loss: 0.05808864161372185
step: 390, loss: 0.02931114286184311
step: 400, loss: 0.10137902200222015
step: 410, loss: 0.0005463444394990802
step: 420, loss: 0.08868152648210526
step: 430, loss: 0.029565270990133286
step: 440, loss: 0.02749599888920784
step: 450, loss: 0.011857527308166027
step: 460, loss: 0.2376682013273239
step: 470, loss: 0.0699046328663826
step: 480, loss: 0.27595484256744385
step: 490, loss: 0.028317907825112343
step: 500, loss: 0.034177105873823166
step: 510, loss: 0.10626853257417679
step: 520, loss: 0.06789112836122513
step: 530, loss: 0.16697555780410767

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

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9299771167048057, f1=0.9271644525881814,
best_f1=0.9271644525881814
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.036016277968883514
step: 10, loss: 0.16113027930259705
step: 20, loss: 0.2070939540863037
step: 30, loss: 0.2862538695335388
step: 40, loss: 0.03145808354020119
step: 50, loss: 0.014822694472968578
step: 60, loss: 0.0063088261522352695
step: 70, loss: 0.037120286375284195
step: 80, loss: 0.04407261312007904
step: 90, loss: 0.03088005632162094
step: 100, loss: 0.06343186646699905
step: 110, loss: 0.05414021015167236
step: 120, loss: 0.09419353306293488
step: 130, loss: 0.017512831836938858
step: 140, loss: 0.016561565920710564
step: 150, loss: 0.01141582801938057
step: 160, loss: 0.014447754248976707
step: 170, loss: 0.011847218498587608
step: 180, loss: 0.10844311118125916
step: 190, loss: 0.015531456097960472
step: 200, loss: 0.017584864050149918
step: 210, loss: 0.09148776531219482
step: 220, loss: 0.011883072555065155
step: 230, loss: 0.02714690938591957
step: 240, loss: 0.0063215517438948154
step: 250, loss: 0.015125117264688015
step: 260, loss: 0.00823171716183424
step: 270, loss: 0.001642898772843182
step: 280, loss: 0.04936725273728371
step: 290, loss: 0.01922641135752201
step: 300, loss: 0.057764094322919846
step: 310, loss: 0.10814928263425827
step: 320, loss: 0.16144026815891266
step: 330, loss: 0.005250217858701944
step: 340, loss: 0.10014398396015167
step: 350, loss: 0.005870320368558168
step: 360, loss: 0.003909004852175713

```

```

step: 370, loss: 0.004694458097219467
step: 380, loss: 0.004799684509634972
step: 390, loss: 0.05596170946955681
step: 400, loss: 0.06427400559186935
step: 410, loss: 0.009507819078862667
step: 420, loss: 0.07645010948181152
step: 430, loss: 0.031022338196635246
step: 440, loss: 0.01270382758229971
step: 450, loss: 0.030255457386374474
step: 460, loss: 0.06233305484056473
step: 470, loss: 0.011614124290645123
step: 480, loss: 0.016444750130176544
step: 490, loss: 0.009459824301302433
step: 500, loss: 0.10969717055559158
step: 510, loss: 0.0011546891182661057
step: 520, loss: 0.013985303230583668
step: 530, loss: 0.0616285540163517
/usr/local/lib/python3.7/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=0.9269406392694064, f1=0.9254686785550983,
best_f1=0.9271644525881814
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06676649302244186
step: 10, loss: 0.0957137793302536
step: 20, loss: 0.11460478603839874
step: 30, loss: 0.03185386583209038
step: 40, loss: 0.0020013751927763224
step: 50, loss: 0.015308576636016369
step: 60, loss: 0.0022430000826716423
step: 70, loss: 0.001949936500750482
step: 80, loss: 0.01190529391169548
step: 90, loss: 0.037445809692144394
step: 100, loss: 0.09753262996673584
step: 110, loss: 0.024195022881031036
step: 120, loss: 0.002906421897932887
step: 130, loss: 0.030011780560016632
step: 140, loss: 0.0343792550265789
step: 150, loss: 0.0013458647299557924
step: 160, loss: 0.006997595075517893
step: 170, loss: 0.0018135138088837266
step: 180, loss: 0.04340007156133652
step: 190, loss: 0.025128519162535667
step: 200, loss: 0.0028755664825439453
step: 210, loss: 0.1045551672577858
step: 220, loss: 0.01494880672544241

```

```

step: 230, loss: 0.21622882783412933
step: 240, loss: 0.03896870091557503
step: 250, loss: 0.09486186504364014
step: 260, loss: 0.15606220066547394
step: 270, loss: 0.05575486272573471
step: 280, loss: 0.010640808381140232
step: 290, loss: 0.05325120687484741
step: 300, loss: 0.022203294560313225
step: 310, loss: 0.004216773901134729
step: 320, loss: 0.034021440893411636
step: 330, loss: 0.01632647030055523
step: 340, loss: 0.0016123513923957944
step: 350, loss: 0.002293981146067381
step: 360, loss: 0.08319839835166931
step: 370, loss: 0.04795307666063309
step: 380, loss: 0.008123228326439857
step: 390, loss: 0.029429184272885323
step: 400, loss: 0.03289837762713432
step: 410, loss: 0.003698749700561166
step: 420, loss: 0.04925169795751572
step: 430, loss: 0.008523315191268921
step: 440, loss: 0.06664945185184479
step: 450, loss: 0.017043545842170715
step: 460, loss: 0.0029784864746034145
step: 470, loss: 0.030711257830262184
step: 480, loss: 0.0007631508633494377
step: 490, loss: 0.011530125513672829
step: 500, loss: 0.00882788747549057
step: 510, loss: 0.05069030076265335
step: 520, loss: 0.13767650723457336
step: 530, loss: 0.008113577030599117
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9317865429234339, f1=0.9285714285714285,
best_f1=0.9285714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012603018432855606
step: 10, loss: 0.013014690019190311
step: 20, loss: 0.008269778452813625
step: 30, loss: 0.0011045774444937706
step: 40, loss: 0.03472820669412613

```

step: 50, loss: 0.005823073908686638
step: 60, loss: 0.0020157983526587486
step: 70, loss: 0.007819486781954765
step: 80, loss: 0.001060717971995473
step: 90, loss: 0.19573451578617096
step: 100, loss: 0.005479960236698389
step: 110, loss: 0.0033631629776209593
step: 120, loss: 0.004621181171387434
step: 130, loss: 0.004363603889942169
step: 140, loss: 0.0009099068120121956
step: 150, loss: 0.0024294378235936165
step: 160, loss: 0.007308891974389553
step: 170, loss: 0.0642913281917572
step: 180, loss: 0.004990212619304657
step: 190, loss: 0.0007120973314158618
step: 200, loss: 0.003776563797146082
step: 210, loss: 0.004898618441075087
step: 220, loss: 0.0019079004414379597
step: 230, loss: 0.013394900597631931
step: 240, loss: 0.0047050463035702705
step: 250, loss: 0.0005384665564633906
step: 260, loss: 0.003245915286242962
step: 270, loss: 0.14545458555221558
step: 280, loss: 0.06257136166095734
step: 290, loss: 0.0016261014388874173
step: 300, loss: 0.08542437106370926
step: 310, loss: 0.000690823420882225
step: 320, loss: 0.02465079538524151
step: 330, loss: 0.001015376765280962
step: 340, loss: 0.05682697147130966
step: 350, loss: 0.003544565523043275
step: 360, loss: 0.003951712045818567
step: 370, loss: 0.0005233159172348678
step: 380, loss: 0.08609431236982346
step: 390, loss: 0.007937177084386349
step: 400, loss: 0.03899853304028511
step: 410, loss: 0.09767816960811615
step: 420, loss: 0.005235569551587105
step: 430, loss: 0.014874167740345001
step: 440, loss: 0.16047967970371246
step: 450, loss: 0.1989866942167282
step: 460, loss: 0.0044479528442025185
step: 470, loss: 0.0110256876796484
step: 480, loss: 0.0033541149459779263
step: 490, loss: 0.008777085691690445
step: 500, loss: 0.0012927058851346374
step: 510, loss: 0.0080221276730299
step: 520, loss: 0.11845459789037704

```
step: 530, loss: 0.001012568362057209
/usr/local/lib/python3.7/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=0.9260273972602739, f1=0.9201101928374655,
best_f1=0.9285714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012838783673942089
step: 10, loss: 0.0027382189873605967
step: 20, loss: 0.006095917429775
step: 30, loss: 0.0028047228697687387
step: 40, loss: 0.004542966838926077
step: 50, loss: 0.07043708115816116
step: 60, loss: 0.0022287468891590834
step: 70, loss: 0.029836343601346016
step: 80, loss: 0.0015460121212527156
step: 90, loss: 0.0012449492933228612
step: 100, loss: 0.07488633692264557
step: 110, loss: 0.0077023981139063835
step: 120, loss: 0.020752718672156334
step: 130, loss: 0.0014761447673663497
step: 140, loss: 0.0007955962209962308
step: 150, loss: 0.001487534842453897
step: 160, loss: 0.0011086612939834595
step: 170, loss: 0.00048611281090416014
step: 180, loss: 0.003527645021677017
step: 190, loss: 0.033634357154369354
step: 200, loss: 0.0007719340501353145
step: 210, loss: 0.0014329812256619334
step: 220, loss: 0.0029312961269170046
step: 230, loss: 0.00016783819592092186
step: 240, loss: 0.00032600024132989347
step: 250, loss: 0.00016181296086870134
step: 260, loss: 0.08115185797214508
step: 270, loss: 0.00032400520285591483
step: 280, loss: 0.001694236765615642
step: 290, loss: 0.00790504552423954
step: 300, loss: 0.002765769138932228
step: 310, loss: 0.00039253875729627907
step: 320, loss: 0.03283097594976425
step: 330, loss: 0.012133928947150707
step: 340, loss: 0.010003648698329926
step: 350, loss: 0.13782340288162231
step: 360, loss: 0.012115344405174255
step: 370, loss: 0.010418584570288658
step: 380, loss: 0.0032359107863157988
```



```

step: 390, loss: 0.006150877568870783
step: 400, loss: 0.00027228580438531935
step: 410, loss: 0.00013565261906478554
step: 420, loss: 0.00822339579463005
step: 430, loss: 0.02130233123898506
step: 440, loss: 0.07058502733707428
step: 450, loss: 0.0010258349357172847
step: 460, loss: 0.00596276018768549
step: 470, loss: 0.06493514776229858
step: 480, loss: 0.006055950187146664
step: 490, loss: 0.0019510603742673993
step: 500, loss: 0.0013399564195424318
step: 510, loss: 0.00043412199011072516
step: 520, loss: 0.007425589952617884
step: 530, loss: 0.003485833527520299
/usr/local/lib/python3.7/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=0.9212121212121211, f1=0.9183673469387754,
best_f1=0.9285714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0020734593272209167
step: 10, loss: 0.009626932442188263
step: 20, loss: 0.025074604898691177
step: 30, loss: 0.008636802434921265
step: 40, loss: 0.004220947623252869
step: 50, loss: 0.0026481598615646362
step: 60, loss: 0.0030747130513191223
step: 70, loss: 0.001431780750863254
step: 80, loss: 0.01839965581893921
step: 90, loss: 0.0004559574299491942
step: 100, loss: 0.0018847070168703794
step: 110, loss: 0.0017647516215220094
step: 120, loss: 0.0020794605370610952
step: 130, loss: 0.00038614278309978545
step: 140, loss: 0.004665196407586336
step: 150, loss: 0.003646235913038254
step: 160, loss: 0.015114146284759045
step: 170, loss: 0.039004817605018616
step: 180, loss: 0.0008862669346854091
step: 190, loss: 0.00021129456581547856
step: 200, loss: 0.012901248410344124
step: 210, loss: 0.0016837141010910273
step: 220, loss: 0.0003497749858070165
step: 230, loss: 0.000341676059179008
step: 240, loss: 0.002781973220407963

```

```

step: 250, loss: 0.004289192613214254
step: 260, loss: 0.0038434152957051992
step: 270, loss: 0.08087294548749924
step: 280, loss: 0.07648646086454391
step: 290, loss: 0.0032082111574709415
step: 300, loss: 0.0005422181566245854
step: 310, loss: 0.000885715417098254
step: 320, loss: 0.035018086433410645
step: 330, loss: 0.0004850133555009961
step: 340, loss: 0.023233965039253235
step: 350, loss: 0.0020357253961265087
step: 360, loss: 0.009362607263028622
step: 370, loss: 0.06976000964641571
step: 380, loss: 0.0009903482859954238
step: 390, loss: 7.237918907776475e-05
step: 400, loss: 8.67346316226758e-05
step: 410, loss: 0.001448378199711442
step: 420, loss: 0.0017804332310333848
step: 430, loss: 0.00010824044147739187
step: 440, loss: 0.011096254922449589
step: 450, loss: 0.005979835521429777
step: 460, loss: 0.0035238447599112988
step: 470, loss: 0.04654858633875847
step: 480, loss: 0.0004495530156418681
step: 490, loss: 0.0013196839718148112
step: 500, loss: 0.0003010295331478119
step: 510, loss: 0.022128891199827194
step: 520, loss: 0.006048878189176321
step: 530, loss: 0.00045674864668399096
/usr/local/lib/python3.7/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=0.9258753979081401, f1=0.9245541838134431,
best_f1=0.9285714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010441886261105537
step: 10, loss: 0.0041960496455430984
step: 20, loss: 0.006092366762459278
step: 30, loss: 0.068399578332901
step: 40, loss: 0.00030109466752037406
step: 50, loss: 0.0014853281900286674
step: 60, loss: 0.023737290874123573
step: 70, loss: 0.007250974420458078
step: 80, loss: 0.00019216105283703655
step: 90, loss: 0.00798803847283125
step: 100, loss: 0.008111434057354927

```

```
step: 110, loss: 0.002172174397855997
step: 120, loss: 0.00681393314152956
step: 130, loss: 0.06319140642881393
step: 140, loss: 0.006682527717202902
step: 150, loss: 0.00688360957428813
step: 160, loss: 0.012270445004105568
step: 170, loss: 0.0019504022784531116
step: 180, loss: 0.00297447107732296
step: 190, loss: 0.0003341846168041229
step: 200, loss: 0.0007285543251782656
step: 210, loss: 0.006214916706085205
step: 220, loss: 0.008417966775596142
step: 230, loss: 8.981449354905635e-05
step: 240, loss: 8.328879630425945e-05
step: 250, loss: 0.005878030322492123
step: 260, loss: 0.029599154368042946
step: 270, loss: 6.196632602950558e-05
step: 280, loss: 0.00028177734930068254
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 290, loss: 0.019823914393782616
step: 300, loss: 0.0002308265247847885
step: 310, loss: 0.031073281541466713
step: 320, loss: 0.0016082135261967778
step: 330, loss: 0.03493395820260048
step: 340, loss: 0.0023522633127868176
step: 350, loss: 0.0012851713690906763
step: 360, loss: 0.0002779969363473356
step: 370, loss: 0.0005112735670991242
step: 380, loss: 0.01787680760025978
step: 390, loss: 0.008972711861133575
step: 400, loss: 0.015534414909780025
step: 410, loss: 0.0005076027591712773
step: 420, loss: 0.00036736129550263286
step: 430, loss: 0.0007603827398270369
step: 440, loss: 0.001826070831157267
step: 450, loss: 0.0002893274649977684
step: 460, loss: 0.0010557951172813773
step: 470, loss: 0.017734425142407417
step: 480, loss: 0.00013030013360548764
step: 490, loss: 0.0007890411652624607
step: 500, loss: 0.0015147473895922303
step: 510, loss: 0.001447970629669726
step: 520, loss: 0.007525305263698101
step: 530, loss: 0.0010112767340615392
/usr/local/lib/python3.7/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=0.9213911386374464, f1=0.9154795821462489,
```

```
best_f1=0.9285714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00516149029135704
step: 10, loss: 0.0016001404728740454
step: 20, loss: 0.0003305970458313823
step: 30, loss: 0.015225637704133987
step: 40, loss: 0.011307025328278542
step: 50, loss: 0.00012272322783246636
step: 60, loss: 0.009918772615492344
step: 70, loss: 0.12418079376220703
step: 80, loss: 0.00047924055252224207
step: 90, loss: 0.013833981938660145
step: 100, loss: 0.00010312319500371814
step: 110, loss: 0.011138821952044964
step: 120, loss: 0.004014324862509966
step: 130, loss: 0.00020154181402176619
step: 140, loss: 0.17395097017288208
step: 150, loss: 0.0015947532374411821
step: 160, loss: 0.0004711448273155838
step: 170, loss: 0.0008379669161513448
step: 180, loss: 0.00022756843827664852
step: 190, loss: 0.008834977634251118
step: 200, loss: 0.00021110243687871844
step: 210, loss: 0.0008625454502180219
step: 220, loss: 0.0015186585951596498
step: 230, loss: 0.0003194573801010847
step: 240, loss: 0.0003208029374945909
step: 250, loss: 0.000992055400274694
step: 260, loss: 0.011299577541649342
step: 270, loss: 0.06838028877973557
step: 280, loss: 9.07771463971585e-05
step: 290, loss: 0.00026009423891082406
step: 300, loss: 0.00031266859150491655
step: 310, loss: 0.0003744007262866944
step: 320, loss: 0.0028294608928263187
step: 330, loss: 0.0010562674142420292
step: 340, loss: 0.0023892768658697605
step: 350, loss: 0.0004810439713764936
step: 360, loss: 0.02206047624349594
step: 370, loss: 0.001439403393305838
step: 380, loss: 0.00022741650172974914
step: 390, loss: 0.0018981850007548928
step: 400, loss: 0.05484339967370033
step: 410, loss: 0.0003661883529275656
step: 420, loss: 0.0032336306758224964
step: 430, loss: 0.00010607122385408729
```

```

step: 440, loss: 0.000326104142004624
step: 450, loss: 0.0010286073666065931
step: 460, loss: 0.00108368752989918
step: 470, loss: 0.0006718820077367127
step: 480, loss: 9.251909796148539e-05
step: 490, loss: 0.0002330987626919523
step: 500, loss: 0.006407599896192551
step: 510, loss: 0.020246323198080063
step: 520, loss: 0.0001527033600723371
step: 530, loss: 0.00014851668674964458
/usr/local/lib/python3.7/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=0.9243140964995269, f1=0.9254716981132075,
best_f1=0.9285714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003732088254764676
step: 10, loss: 0.0002590582880657166
step: 20, loss: 0.0004078999045304954
step: 30, loss: 0.0009887078776955605
step: 40, loss: 0.0011653152760118246
step: 50, loss: 0.000185557670192793
step: 60, loss: 0.00013791759556625038
step: 70, loss: 0.002265369053930044
step: 80, loss: 0.003301797667518258
step: 90, loss: 0.06962660700082779
step: 100, loss: 8.976605022326112e-05
step: 110, loss: 0.0012217991752550006
step: 120, loss: 0.0018130001844838262
step: 130, loss: 0.0011150323553010821
step: 140, loss: 0.004542430397123098
step: 150, loss: 0.00014509724860545248
step: 160, loss: 0.0022406610660254955
step: 170, loss: 0.0024422514252364635
step: 180, loss: 0.001471653231419623
step: 190, loss: 0.00033910272759385407
step: 200, loss: 0.0008793892338871956
step: 210, loss: 9.635043534217402e-05
step: 220, loss: 0.006444482132792473
step: 230, loss: 0.00014019606169313192
step: 240, loss: 6.282816320890561e-05
step: 250, loss: 0.0021594802383333445
step: 260, loss: 0.0005759703926742077
step: 270, loss: 7.896437455201522e-05
step: 280, loss: 0.03327646479010582
step: 290, loss: 0.0013616771902889013

```

```

step: 300, loss: 0.002529314486309886
step: 310, loss: 0.0009539728634990752
step: 320, loss: 0.0009174350998364389
step: 330, loss: 0.00977400690317154
step: 340, loss: 7.222894782898948e-05
step: 350, loss: 0.00019761850126087666
step: 360, loss: 0.03586664795875549
step: 370, loss: 0.014350666664540768
step: 380, loss: 0.0014500089455395937
step: 390, loss: 0.00032119647949002683
step: 400, loss: 0.00041137717198580503
step: 410, loss: 4.010892371297814e-05
step: 420, loss: 0.03177327662706375
step: 430, loss: 0.00015995721332728863
step: 440, loss: 0.000566454604268074
step: 450, loss: 0.0009446185431443155
step: 460, loss: 0.0034466406796127558
step: 470, loss: 0.00056277314433828
step: 480, loss: 0.0030238088220357895
step: 490, loss: 0.06588728725910187
step: 500, loss: 0.044214896857738495
step: 510, loss: 0.002068879548460245
step: 520, loss: 0.0020098029635846615
step: 530, loss: 0.0002095360541716218
/usr/local/lib/python3.7/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=0.9278445883441258, f1=0.9314179796107508,
best_f1=0.9285714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002278005238622427
step: 10, loss: 0.0033724268432706594
step: 20, loss: 7.550571899628267e-05
step: 30, loss: 0.015368575230240822
step: 40, loss: 0.0006979090394452214
step: 50, loss: 0.002298827050253749
step: 60, loss: 0.0008377643534913659
step: 70, loss: 0.0002258180029457435
step: 80, loss: 0.014577565714716911
step: 90, loss: 0.001365660224109888
step: 100, loss: 0.00021994870621711016
step: 110, loss: 0.000568955612834543
step: 120, loss: 0.06511431187391281
step: 130, loss: 0.07555787265300751
step: 140, loss: 0.001363176736049354
step: 150, loss: 0.010266476310789585

```

```

step: 160, loss: 0.0038233520463109016
step: 170, loss: 0.001991486642509699
step: 180, loss: 6.500392191810533e-05
step: 190, loss: 0.0010436446173116565
step: 200, loss: 0.0011319989571347833
step: 210, loss: 0.0019116170005872846
step: 220, loss: 0.0006262360257096589
step: 230, loss: 0.0325637012720108
step: 240, loss: 0.0026772734709084034
step: 250, loss: 8.274262654595077e-05
step: 260, loss: 5.334385059541091e-05
step: 270, loss: 0.0020806966349482536
step: 280, loss: 0.003928366582840681
step: 290, loss: 0.0019636524375528097
step: 300, loss: 0.013916303403675556
step: 310, loss: 0.0006767387967556715
step: 320, loss: 0.015253693796694279
step: 330, loss: 0.0007176348008215427
step: 340, loss: 0.00012187231186544523
step: 350, loss: 0.0029977471567690372
step: 360, loss: 0.019427701830863953
step: 370, loss: 7.248236215673387e-05
step: 380, loss: 4.943756721331738e-05
step: 390, loss: 0.03667139261960983
step: 400, loss: 4.459293995751068e-05
step: 410, loss: 0.002180437557399273
step: 420, loss: 0.0007412669365294278
step: 430, loss: 0.014116709120571613
step: 440, loss: 0.0045665232464671135
step: 450, loss: 0.0017251267563551664
step: 460, loss: 0.0153540950268507
step: 470, loss: 0.022120220586657524
step: 480, loss: 0.00036009689210914075
step: 490, loss: 0.009009003639221191
step: 500, loss: 0.0005866211140528321
step: 510, loss: 5.3807318181497976e-05
step: 520, loss: 0.0002834126935340464
step: 530, loss: 0.00013634182687383145
/usr/local/lib/python3.7/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=0.9250116441546343, f1=0.9201127819548872,
best_f1=0.9285714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00045164325274527073
step: 10, loss: 0.0792803093791008

```

step: 20, loss: 4.2579245928209275e-05
step: 30, loss: 0.0004623365239240229
step: 40, loss: 0.0025384249165654182
step: 50, loss: 0.002059880644083023
step: 60, loss: 0.0003587533428799361
step: 70, loss: 0.00022822873143013567
step: 80, loss: 0.0017612949013710022
step: 90, loss: 0.08436625450849533
step: 100, loss: 0.0009696225170046091
step: 110, loss: 0.0003763861604966223
step: 120, loss: 0.000139221825520508
step: 130, loss: 0.0001404750073561445
step: 140, loss: 0.003238291945308447
step: 150, loss: 0.0016271042404696345
step: 160, loss: 0.00048384323599748313
step: 170, loss: 7.342669414356351e-05
step: 180, loss: 0.013752064667642117
step: 190, loss: 0.0002146384067600593
step: 200, loss: 0.00013247155584394932
step: 210, loss: 0.0013079550117254257
step: 220, loss: 5.02009570482187e-05
step: 230, loss: 0.000773095351178199
step: 240, loss: 0.0007835480500943959
step: 250, loss: 8.199881995096803e-05
step: 260, loss: 0.0005631559761241078
step: 270, loss: 0.00031335162930190563
step: 280, loss: 0.0008235137211158872
step: 290, loss: 0.00037724507274106145
step: 300, loss: 7.680562703171745e-05
step: 310, loss: 0.003934631124138832
step: 320, loss: 0.0002258450404042378
step: 330, loss: 0.00014143239241093397
step: 340, loss: 0.0007201298722065985
step: 350, loss: 0.042023640125989914
step: 360, loss: 0.02478039637207985
step: 370, loss: 0.00016939638589974493
step: 380, loss: 0.0015759117668494582
step: 390, loss: 0.18539580702781677
step: 400, loss: 0.020128196105360985
step: 410, loss: 0.0008392130839638412
step: 420, loss: 0.0005194672849029303
step: 430, loss: 0.0002974089584313333
step: 440, loss: 0.0021261991932988167
step: 450, loss: 0.07088010013103485
step: 460, loss: 0.0013884035870432854
step: 470, loss: 0.00022008424275554717
step: 480, loss: 0.000146931255585514
step: 490, loss: 0.0018145940266549587


```

step: 500, loss: 0.0013158622896298766
step: 510, loss: 6.246184784686193e-05
step: 520, loss: 0.0016340777510777116
step: 530, loss: 0.00016784125182311982
/usr/local/lib/python3.7/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=0.9267840593141797, f1=0.926012098650535,
best_f1=0.9285714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02605733647942543
step: 10, loss: 0.00022171481396071613
step: 20, loss: 0.01585068739950657
step: 30, loss: 0.0357406847178936
step: 40, loss: 0.00010365369962528348
step: 50, loss: 0.00012521271128207445
step: 60, loss: 0.005135908257216215
step: 70, loss: 0.0015596161829307675
step: 80, loss: 3.829749766737223e-05
step: 90, loss: 0.0003894676628988236
step: 100, loss: 0.0016904058866202831
step: 110, loss: 0.0001449580304324627
step: 120, loss: 0.00013344686885830015
step: 130, loss: 0.00010762042802525684
step: 140, loss: 0.00018231519788969308
step: 150, loss: 0.00018077075947076082
step: 160, loss: 3.089244637521915e-05
step: 170, loss: 0.0004796223947778344
step: 180, loss: 0.0001006243473966606
step: 190, loss: 0.00013424963981378824
step: 200, loss: 0.009106220677495003
step: 210, loss: 0.0025625445414334536
step: 220, loss: 0.0006508283549919724
step: 230, loss: 0.003572509391233325
step: 240, loss: 5.715028601116501e-05
step: 250, loss: 0.030970362946391106
step: 260, loss: 0.0008484856225550175
step: 270, loss: 0.0018957291031256318
step: 280, loss: 0.0012174673611298203
step: 290, loss: 0.00042912515345960855
step: 300, loss: 0.0008501614793203771
step: 310, loss: 0.005281352903693914
step: 320, loss: 0.013265401124954224
step: 330, loss: 0.0016570863081142306
step: 340, loss: 1.9009861716767773e-05
step: 350, loss: 0.009372308850288391

```

```

step: 360, loss: 2.5643939807196148e-05
step: 370, loss: 0.001661204849369824
step: 380, loss: 0.0001336675341008231
step: 390, loss: 0.0006850074860267341
step: 400, loss: 3.9782658859621733e-05
step: 410, loss: 4.351207098807208e-05
step: 420, loss: 0.00010516739712329581
step: 430, loss: 0.06567174196243286
step: 440, loss: 0.00015107270155567676
step: 450, loss: 5.908786260988563e-05
step: 460, loss: 1.832066300266888e-05
step: 470, loss: 0.0143464095890522
step: 480, loss: 0.001263176091015339
step: 490, loss: 9.467080235481262e-05
step: 500, loss: 0.00013421379844658077
step: 510, loss: 0.001579443458467722
step: 520, loss: 0.00013670913176611066
step: 530, loss: 3.9527436456410214e-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=0.9273323956868261, f1=0.9240150093808631,
best_f1=0.9285714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.050173684954643e-05
step: 10, loss: 0.0019764022435992956
step: 20, loss: 0.0001546147686894983
step: 30, loss: 0.0003091166145168245
step: 40, loss: 4.169575186097063e-05
step: 50, loss: 0.0030458588153123856
step: 60, loss: 3.278447547927499e-05
step: 70, loss: 0.00021088149514980614
step: 80, loss: 0.0004778109141625464
step: 90, loss: 9.19731828616932e-05
step: 100, loss: 0.000310762959998101
step: 110, loss: 0.00014134999946691096
step: 120, loss: 0.00010038838809123263
step: 130, loss: 0.0006464031175710261
step: 140, loss: 0.0056652468629181385
step: 150, loss: 0.00029734676354564726
step: 160, loss: 0.001218305085785687
step: 170, loss: 0.0089223962277174
step: 180, loss: 0.0014021525857970119
step: 190, loss: 0.0001581784599693492
step: 200, loss: 0.0010631955228745937
step: 210, loss: 0.03788481280207634

```

```

step: 220, loss: 3.679474684759043e-05
step: 230, loss: 0.0008503354620188475
step: 240, loss: 0.0005571607034653425
step: 250, loss: 1.5240021639328916e-05
step: 260, loss: 1.9311548385303468e-05
step: 270, loss: 0.0023340261541306973
step: 280, loss: 0.0001786422508303076
step: 290, loss: 0.00021058043057564646
step: 300, loss: 0.00022987538250163198
step: 310, loss: 2.3040396627038717e-05
step: 320, loss: 0.0008916176157072186
step: 330, loss: 0.00010510421998333186
step: 340, loss: 4.981396341463551e-05
step: 350, loss: 0.0012047591153532267
step: 360, loss: 0.05085797235369682
step: 370, loss: 9.417504770681262e-05
step: 380, loss: 0.00014758225006517023
step: 390, loss: 0.003684087423607707
step: 400, loss: 2.9591208658530377e-05
step: 410, loss: 0.0013987721176818013
step: 420, loss: 0.00037807837361469865
step: 430, loss: 0.00011445055861258879
step: 440, loss: 2.523786315578036e-05
step: 450, loss: 0.0003616652684286237
step: 460, loss: 0.00016114335448946804
step: 470, loss: 0.00026783495559357107
step: 480, loss: 6.11097930232063e-05
step: 490, loss: 3.832291258731857e-05
step: 500, loss: 0.03302307799458504
step: 510, loss: 0.00021279750217217952
step: 520, loss: 5.5407243053196e-05
step: 530, loss: 9.778487583389506e-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=0.9251510925151093, f1=0.927806241266884,
best_f1=0.9285714285714285
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.595468297135085e-05
step: 10, loss: 3.489635491860099e-05
step: 20, loss: 4.6651228331029415e-05
step: 30, loss: 5.738232721341774e-05
step: 40, loss: 0.00022763971355743706
step: 50, loss: 0.0002724523073993623
step: 60, loss: 0.00013420952018350363
step: 70, loss: 0.00019765165052376688

```

```
step: 80, loss: 9.057468560058624e-05
step: 90, loss: 1.770984636095818e-05
step: 100, loss: 0.0011054239002987742
step: 110, loss: 0.004264806397259235
step: 120, loss: 0.0017840773798525333
step: 130, loss: 5.9941612562397495e-05
step: 140, loss: 8.710815018275753e-05
step: 150, loss: 9.045363549375907e-05
step: 160, loss: 3.703883703565225e-05
step: 170, loss: 0.00010627614392433316
step: 180, loss: 3.066411954932846e-05
step: 190, loss: 0.0005518443649634719
step: 200, loss: 0.0014894896885380149
step: 210, loss: 6.789130566176027e-05
step: 220, loss: 0.00012006665929220617
step: 230, loss: 0.10044408589601517
step: 240, loss: 0.0006133728311397135
step: 250, loss: 6.88844666001387e-05
step: 260, loss: 0.0009376351372338831
step: 270, loss: 0.010064022615551949
step: 280, loss: 0.00028664589626714587
step: 290, loss: 0.0007227088790386915
step: 300, loss: 0.0013651956105604768
step: 310, loss: 0.000848367577418685
step: 320, loss: 0.0008721020421944559
step: 330, loss: 0.001562082557938993
step: 340, loss: 0.00017157025285996497
step: 350, loss: 0.0005642544128932059
step: 360, loss: 0.00041929344297386706
step: 370, loss: 0.003747215960174799
step: 380, loss: 5.429784141597338e-05
step: 390, loss: 2.6969370082952082e-05
step: 400, loss: 0.00044343905756250024
step: 410, loss: 0.000495004584081471
step: 420, loss: 0.0004113295581191778
step: 430, loss: 6.951782415853813e-05
step: 440, loss: 0.001285316189751029
step: 450, loss: 0.0016999810468405485
step: 460, loss: 0.00018953489779960364
step: 470, loss: 0.0001680427376413718
step: 480, loss: 0.0020898908842355013
step: 490, loss: 0.0004990058951079845
step: 500, loss: 7.831695984350517e-05
step: 510, loss: 9.999750909628347e-05
step: 520, loss: 0.00036104791797697544
step: 530, loss: 0.00013461401977110654
/usr/local/lib/python3.7/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=0.927710843373494, f1=0.9285051067780873,
best_f1=0.9285714285714285
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, 357.14it/s]
load_f1 = 0.9319664492078286
real_f1 = 0.9288040949278734
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, 359.41it/s]

```

2 BASELINE

2.1 BASELINE STRUCTURED

2.1.1 Beer - Running the matcher

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

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

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

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

Defaults for this optimization level are:

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

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

```

After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8540323376655579
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.3684210526315789, f1=0.2978723404255319,
best_f1=0.2978723404255319
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3612845540046692
/usr/local/lib/python3.7/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=0.3448275862068965, f1=0.32911392405063294,
best_f1=0.2978723404255319
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.32213786244392395
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.3835616438356164, f1=0.2978723404255319,
best_f1=0.2978723404255319
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.32970893383026123
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.39285714285714285, f1=0.33766233766233766,
best_f1=0.33766233766233766
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.287137895822525
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.42857142857142855, f1=0.32, best_f1=0.32
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2242448329925537
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5, f1=0.4, best_f1=0.4
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```



```

step: 0, loss: 0.1896531730890274
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 7: dev_f1=0.5263157894736842, f1=0.3492063492063492,
best_f1=0.3492063492063492
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.294527143239975
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5454545454545454, f1=0.3728813559322034,
best_f1=0.3728813559322034
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.18600508570671082
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5833333333333334, f1=0.42553191489361697,
best_f1=0.42553191489361697
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.16427592933177948
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6060606060606061, f1=0.3728813559322034,
best_f1=0.3728813559322034
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.13836701214313507
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6153846153846153, f1=0.4, best_f1=0.4
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08411098271608353
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6250000000000001, f1=0.423076923076923,
best_f1=0.423076923076923
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.057337235659360886
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6451612903225806, f1=0.43137254901960786,
best_f1=0.43137254901960786
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10486294329166412
/usr/local/lib/python3.7/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=0.6451612903225806, f1=0.43137254901960786,

```

```

best_f1=0.43137254901960786
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.15961936116218567
/usr/local/lib/python3.7/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=0.6451612903225806, f1=0.43137254901960786,
best_f1=0.43137254901960786
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
91it [00:00, 105728.99it/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 = 0.6451612903225806
real_f1 = 0.6111111111111112
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, 256.83it/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: 0.8127899169921875
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4816610515117645
step: 20, loss: 0.5845920443534851
step: 30, loss: 0.4723525941371918
step: 40, loss: 0.2849979102611542
step: 50, loss: 0.09906496107578278
step: 60, loss: 0.0807574912905693
step: 70, loss: 0.20777980983257294
step: 80, loss: 0.1657734364271164
step: 90, loss: 0.028934111818671227
step: 100, loss: 0.0908610001206398
step: 110, loss: 0.03589678555727005
step: 120, loss: 0.07422485947608948
step: 130, loss: 0.004792419727891684
step: 140, loss: 0.015221435576677322
step: 150, loss: 0.07153996080160141
step: 160, loss: 0.16696913540363312
step: 170, loss: 0.008871105499565601
step: 180, loss: 0.023004507645964622
step: 190, loss: 0.0127872908487916
step: 200, loss: 0.007803497835993767
step: 210, loss: 0.003562267404049635
step: 220, loss: 0.011759012937545776
step: 230, loss: 0.10018706321716309
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:

```

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=0.9775280898876404, f1=0.9683257918552037,

best_f1=0.9683257918552037

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

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

step: 0, loss: 0.00628586346283555

step: 10, loss: 0.001364559167996049

step: 20, loss: 0.0031290962360799313

step: 30, loss: 0.09571422636508942

step: 40, loss: 0.006064277142286301

step: 50, loss: 0.042468663305044174

step: 60, loss: 0.012229298241436481

step: 70, loss: 0.0922604501247406

step: 80, loss: 0.0020666886121034622

step: 90, loss: 0.06142362207174301

step: 100, loss: 0.006367753725498915

step: 110, loss: 0.003584746504202485

step: 120, loss: 0.11315078288316727

step: 130, loss: 0.0030372508335858583

step: 140, loss: 0.027937309816479683

step: 150, loss: 0.021885838359594345

step: 160, loss: 0.020432505756616592

step: 170, loss: 0.07812429219484329

step: 180, loss: 0.009913068264722824

step: 190, loss: 0.021956270560622215

step: 200, loss: 0.0016338981222361326

step: 210, loss: 0.04908891022205353

step: 220, loss: 0.0006859780405648053

step: 230, loss: 0.07862349599599838

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

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

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

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=0.9807037457434733, f1=0.9690721649484535,

best_f1=0.9690721649484535

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

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

step: 0, loss: 0.048859160393476486

step: 10, loss: 0.09237920492887497

step: 20, loss: 0.006620466243475676

step: 30, loss: 0.009876139461994171

```

step: 40, loss: 0.014466777443885803
step: 50, loss: 0.007068203762173653
step: 60, loss: 0.0005072873900644481
step: 70, loss: 0.0007588122389279306
step: 80, loss: 0.018886463716626167
step: 90, loss: 0.0010086552938446403
step: 100, loss: 0.0008730558911338449
step: 110, loss: 0.09665506333112717
step: 120, loss: 0.0017866210546344519
step: 130, loss: 0.0005119714769534767
step: 140, loss: 0.003711757715791464
step: 150, loss: 0.009748848155140877
step: 160, loss: 0.011296925134956837
step: 170, loss: 0.0023879241198301315
step: 180, loss: 0.007790947798639536
step: 190, loss: 0.007483521476387978
step: 200, loss: 0.0018681281944736838
step: 210, loss: 0.024023475125432014
step: 220, loss: 0.002918760059401393
step: 230, loss: 0.009188147261738777
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9887640449438202, f1=0.9797752808988766,
best_f1=0.9797752808988766
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00198310031555593
step: 10, loss: 0.000619857688434422
step: 20, loss: 0.0030779330991208553
step: 30, loss: 0.0011200174922123551
step: 40, loss: 0.006382511928677559
step: 50, loss: 0.0029182806611061096
step: 60, loss: 0.0010979549260810018
step: 70, loss: 0.08050655573606491
step: 80, loss: 0.019854318350553513
step: 90, loss: 0.000898751022759825
step: 100, loss: 0.02089739218354225
step: 110, loss: 0.004998117219656706
step: 120, loss: 0.00553517322987318
step: 130, loss: 0.0012613852741196752
step: 140, loss: 0.0011277429293841124
step: 150, loss: 0.0006564132054336369

```

```

step: 160, loss: 0.000473657448310405
step: 170, loss: 0.0280685443431139
step: 180, loss: 0.042285241186618805
step: 190, loss: 0.005250934977084398
step: 200, loss: 0.0005191022646613419
step: 210, loss: 0.0001619780668988824
step: 220, loss: 0.00038387050153687596
step: 230, loss: 0.000251247372943908
/usr/local/lib/python3.7/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=0.9887387387387387, f1=0.9831271091113611,
best_f1=0.9797752808988766
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00026652036467567086
step: 10, loss: 0.00031417611171491444
step: 20, loss: 0.00025862862821668386
step: 30, loss: 0.00026669877115637064
step: 40, loss: 0.003269472625106573
step: 50, loss: 0.000508470693603158
step: 60, loss: 0.00020439879153855145
step: 70, loss: 0.0012343296548351645
step: 80, loss: 0.00028578893397934735
step: 90, loss: 0.0028272836934775114
step: 100, loss: 0.00806436873972416
step: 110, loss: 0.0003095890278927982
step: 120, loss: 0.028481226414442062
step: 130, loss: 0.11001387238502502
step: 140, loss: 0.0002469935279805213
step: 150, loss: 0.0007047971012070775
step: 160, loss: 0.18649014830589294
step: 170, loss: 0.28747642040252686
step: 180, loss: 0.0009833215735852718
step: 190, loss: 0.0004725500475615263
step: 200, loss: 0.014571926556527615
step: 210, loss: 0.00144485745113343
step: 220, loss: 0.0008841976523399353
step: 230, loss: 0.0035159874241799116
/usr/local/lib/python3.7/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=0.9833147942157954, f1=0.9788182831661093,
best_f1=0.9797752808988766
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```



```

step: 0, loss: 0.03378654271364212
step: 10, loss: 0.0032750493846833706
step: 20, loss: 0.00035685140755958855
step: 30, loss: 0.00036532996455207467
step: 40, loss: 0.0047137257643043995
step: 50, loss: 0.010866605676710606
step: 60, loss: 0.01910630613565445
step: 70, loss: 0.00033698842162266374
step: 80, loss: 0.0029251137748360634
step: 90, loss: 0.00020128232426941395
step: 100, loss: 0.0003459600266069174
step: 110, loss: 0.06605891138315201
step: 120, loss: 0.0004044982197228819
step: 130, loss: 0.013919697143137455
step: 140, loss: 0.00044058775529265404
step: 150, loss: 0.00024332423345185816
step: 160, loss: 0.00028020606259815395
step: 170, loss: 0.00036808312870562077
step: 180, loss: 0.0002768837148323655
step: 190, loss: 0.0002924914297182113
step: 200, loss: 0.00032670257496647537
step: 210, loss: 0.00020675972336903214
step: 220, loss: 0.016379911452531815
step: 230, loss: 0.0001339450536761433
/usr/local/lib/python3.7/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=0.9842696629213483, f1=0.9809203142536477,
best_f1=0.9797752808988766
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02884354256093502
step: 10, loss: 7.509130955440924e-05
step: 20, loss: 0.00021655048476532102
step: 30, loss: 0.0001411601697327569
step: 40, loss: 0.006987608037889004
step: 50, loss: 0.0008060313411988318
step: 60, loss: 0.00029387243557721376
step: 70, loss: 0.006007737945765257
step: 80, loss: 6.374358054017648e-05
step: 90, loss: 0.0009646923281252384
step: 100, loss: 0.006755972281098366
step: 110, loss: 0.0007871366105973721
step: 120, loss: 0.00017139583360403776
step: 130, loss: 0.00029161895508877933
step: 140, loss: 9.385963494423777e-05
step: 150, loss: 0.00012505626364145428

```

```

step: 160, loss: 0.00010323563765268773
step: 170, loss: 0.0041127437725663185
step: 180, loss: 0.00015364112914539874
step: 190, loss: 0.00021692828158847988
step: 200, loss: 0.00040308397728949785
step: 210, loss: 9.84929283731617e-05
step: 220, loss: 0.001559185329824686
step: 230, loss: 0.025691861286759377
/usr/local/lib/python3.7/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=0.9887387387387387, f1=0.9819819819819819,
best_f1=0.9797752808988766
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.023731037974357605
step: 10, loss: 0.018681438639760017
step: 20, loss: 0.0003735914942808449
step: 30, loss: 0.00011672322580125183
step: 40, loss: 9.701481030788273e-05
step: 50, loss: 0.00013244721048977226
step: 60, loss: 0.00021648401161655784
step: 70, loss: 0.00012782917474396527
step: 80, loss: 0.00011290601105429232
step: 90, loss: 7.538516365457326e-05
step: 100, loss: 9.723995754029602e-05
step: 110, loss: 0.00035664564347825944
step: 120, loss: 6.192952423589304e-05
step: 130, loss: 0.0005616305279545486
step: 140, loss: 0.00255834823474288
step: 150, loss: 0.00013188476441428065
step: 160, loss: 9.835160744842142e-05
step: 170, loss: 0.00014046973956283182
step: 180, loss: 0.00091165752382949
step: 190, loss: 0.0023523163981735706
step: 200, loss: 0.0005065355799160898
step: 210, loss: 0.0001313407556153834
step: 220, loss: 0.0002567649644333869
step: 230, loss: 0.01880001276731491
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.990990990990991, f1=0.980963045912654,

```

```

best_f1=0.980963045912654
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00018029780767392367
step: 10, loss: 0.00027476990362629294
step: 20, loss: 0.0024725045077502728
step: 30, loss: 0.0034528477117419243
step: 40, loss: 6.583204958587885e-05
step: 50, loss: 7.577695942018181e-05
step: 60, loss: 0.00017205886251758784
step: 70, loss: 0.00014676613500341773
step: 80, loss: 0.12265828251838684
step: 90, loss: 0.0013370085507631302
step: 100, loss: 0.00039019357063807547
step: 110, loss: 0.0007669345359317958
step: 120, loss: 0.026226727291941643
step: 130, loss: 0.00023396745382342488
step: 140, loss: 0.0002150827640434727
step: 150, loss: 8.672587864566594e-05
step: 160, loss: 0.0002645959903020412
step: 170, loss: 4.9751084588933736e-05
step: 180, loss: 0.00031139806378632784
step: 190, loss: 7.565658597741276e-05
step: 200, loss: 0.00011297105811536312
step: 210, loss: 0.0001613475615158677
step: 220, loss: 0.02642030455172062
step: 230, loss: 0.052492089569568634
/usr/local/lib/python3.7/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=0.9852774631936579, f1=0.9808773903262092,
best_f1=0.980963045912654
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00037602009251713753
step: 10, loss: 0.0007292358786799014
step: 20, loss: 0.0001398010936100036
step: 30, loss: 0.003703773021697998
step: 40, loss: 0.0003218346682842821
step: 50, loss: 0.029941607266664505
step: 60, loss: 0.029752423986792564
step: 70, loss: 0.0012036490952596068
step: 80, loss: 0.0010121308732777834
step: 90, loss: 5.512762436410412e-05
step: 100, loss: 8.75803452800028e-05
step: 110, loss: 6.13791708019562e-05

```

```

step: 120, loss: 0.02108331397175789
step: 130, loss: 0.00010428771929582581
step: 140, loss: 0.011106619611382484
step: 150, loss: 6.43809835310094e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 160, loss: 0.0005740678170695901
step: 170, loss: 0.00017931523325387388
step: 180, loss: 7.655221270397305e-05
step: 190, loss: 6.508755905088037e-05
step: 200, loss: 5.8865876781055704e-05
step: 210, loss: 5.680536196450703e-05
step: 220, loss: 0.012610552832484245
step: 230, loss: 0.0007429139222949743
/usr/local/lib/python3.7/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=0.987598647125141, f1=0.978675645342312,
best_f1=0.980963045912654
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.95210915687494e-05
step: 10, loss: 0.00024204677902162075
step: 20, loss: 6.822354771429673e-05
step: 30, loss: 0.006434333976358175
step: 40, loss: 0.005638490431010723
step: 50, loss: 0.00033937502303160727
step: 60, loss: 6.597633910132572e-05
step: 70, loss: 0.009473851881921291
step: 80, loss: 0.00018251111032441258
step: 90, loss: 0.004778652917593718
step: 100, loss: 0.0002518795372452587
step: 110, loss: 9.890339424600825e-05
step: 120, loss: 0.00028043126803822815
step: 130, loss: 5.415653868112713e-05
step: 140, loss: 5.6657008826732635e-05
step: 150, loss: 5.0447648391127586e-05
step: 160, loss: 0.0009390386985614896
step: 170, loss: 0.021639250218868256
step: 180, loss: 0.00021783055854029953
step: 190, loss: 0.0001459940685890615
step: 200, loss: 6.73645845381543e-05
step: 210, loss: 4.962525417795405e-05
step: 220, loss: 7.446784002240747e-05
step: 230, loss: 0.02228170819580555
/usr/local/lib/python3.7/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=0.9841986455981941, f1=0.9774774774774775,
best_f1=0.980963045912654
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005579298012889922
step: 10, loss: 0.002600989770144224
step: 20, loss: 4.740845906781033e-05
step: 30, loss: 0.00011707590601872653
step: 40, loss: 5.430577584775165e-05
step: 50, loss: 4.2417454096721485e-05
step: 60, loss: 0.012287936173379421
step: 70, loss: 6.3020117522683e-05
step: 80, loss: 9.655872418079525e-05
step: 90, loss: 4.089439971721731e-05
step: 100, loss: 4.126261410419829e-05
step: 110, loss: 5.1592360250651836e-05
step: 120, loss: 5.3756572015117854e-05
step: 130, loss: 6.918044527992606e-05
step: 140, loss: 7.67848250688985e-05
step: 150, loss: 5.2972354751545936e-05
step: 160, loss: 0.030586794018745422
step: 170, loss: 9.470099757891148e-05
step: 180, loss: 6.43259845674038e-05
step: 190, loss: 0.0002818609355017543
step: 200, loss: 0.001092117396183312
step: 210, loss: 5.495879304362461e-05
step: 220, loss: 0.010948114097118378
step: 230, loss: 8.378145139431581e-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=0.9876543209876544, f1=0.9786276715410572,
best_f1=0.980963045912654
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.984136755112559e-05
step: 10, loss: 0.019870257005095482
step: 20, loss: 9.544997737975791e-05
step: 30, loss: 0.022048896178603172
step: 40, loss: 0.0001311129453824833
step: 50, loss: 9.082817268790677e-05
step: 60, loss: 8.316438470501453e-05
step: 70, loss: 5.655255517922342e-05
step: 80, loss: 0.00015654983872082084
step: 90, loss: 0.0024315102491527796
step: 100, loss: 7.508732960559428e-05

```

```

step: 110, loss: 4.950633228872903e-05
step: 120, loss: 4.7360990720335394e-05
step: 130, loss: 6.995419244049117e-05
step: 140, loss: 0.004067606758326292
step: 150, loss: 0.001448743394576013
step: 160, loss: 4.651007475331426e-05
step: 170, loss: 5.68696777918376e-05
step: 180, loss: 0.0021445450838655233
step: 190, loss: 2.991306610056199e-05
step: 200, loss: 4.561627429211512e-05
step: 210, loss: 6.254538311623037e-05
step: 220, loss: 3.331693369545974e-05
step: 230, loss: 3.390985511941835e-05
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 13: dev_f1=0.9887133182844244, f1=0.9819819819819819,
best_f1=0.980963045912654
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.1150804463541135e-05
step: 10, loss: 3.3432595955673605e-05
step: 20, loss: 0.012216540053486824
step: 30, loss: 0.00010975867917295545
step: 40, loss: 3.330994150019251e-05
step: 50, loss: 4.654091753764078e-05
step: 60, loss: 2.605393819976598e-05
step: 70, loss: 0.0006472173263318837
step: 80, loss: 7.046778046060354e-05
step: 90, loss: 9.987771045416594e-05
step: 100, loss: 0.009761904366314411
step: 110, loss: 6.81467863614671e-05
step: 120, loss: 3.2252359233098105e-05
step: 130, loss: 7.655039371456951e-05
step: 140, loss: 4.572080433717929e-05
step: 150, loss: 7.255972741404548e-05
step: 160, loss: 3.0002462153788656e-05
step: 170, loss: 4.576859646476805e-05
step: 180, loss: 6.355840741889551e-05
step: 190, loss: 0.006707821972668171
step: 200, loss: 3.6606445064535365e-05
step: 210, loss: 0.003482429077848792
step: 220, loss: 0.00014790856221225113
step: 230, loss: 2.2835558411316015e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.9887387387387387, f1=0.9796839729119639,
best_f1=0.980963045912654
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.2275416237534955e-05
step: 10, loss: 6.853597733424976e-05
step: 20, loss: 0.002762660151347518
step: 30, loss: 0.00010519829083932564
step: 40, loss: 5.707806485588662e-05
step: 50, loss: 5.4590582294622436e-05
step: 60, loss: 4.3727010051952675e-05
step: 70, loss: 3.961302718380466e-05
step: 80, loss: 0.02332613617181778
step: 90, loss: 2.631852294143755e-05
step: 100, loss: 4.565136259770952e-05
step: 110, loss: 5.405348201747984e-05
step: 120, loss: 0.00013407677761279047
step: 130, loss: 0.00011565770546440035
step: 140, loss: 0.000130428044940345
step: 150, loss: 0.00024378503439947963
step: 160, loss: 5.515011434908956e-05
step: 170, loss: 3.701629611896351e-05
step: 180, loss: 4.15721042372752e-05
step: 190, loss: 0.008382529951632023
step: 200, loss: 3.533653216436505e-05
step: 210, loss: 0.03895098716020584
step: 220, loss: 0.0004945512628182769
step: 230, loss: 9.143923671217635e-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=0.9887133182844244, f1=0.9773755656108598,
best_f1=0.980963045912654
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

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

After processing overrides, optimization options are:

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

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

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

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

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

2473it [00:10, 233.33it/s]

load_f1 = 0.9898762654668166

real_f1 = 0.990990990990991

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, 253.63it/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: 0.8057987689971924

step: 10, loss: 0.41211897134780884
step: 20, loss: 0.4722418189048767
step: 30, loss: 0.4227142632007599
step: 40, loss: 0.3571231961250305
step: 50, loss: 0.1720304787158966
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 60, loss: 0.33642101287841797
step: 70, loss: 0.17388734221458435
step: 80, loss: 0.21530255675315857
step: 90, loss: 0.20642825961112976
step: 100, loss: 0.25529277324676514
step: 110, loss: 0.10232528299093246
step: 120, loss: 0.07782092690467834
step: 130, loss: 0.029673513025045395
step: 140, loss: 0.1723727583885193
step: 150, loss: 0.026705102995038033
step: 160, loss: 0.19236581027507782
step: 170, loss: 0.15617820620536804
step: 180, loss: 0.10782703012228012
step: 190, loss: 0.044692616909742355
step: 200, loss: 0.09024541825056076
step: 210, loss: 0.125576451420784
step: 220, loss: 0.20651613175868988
step: 230, loss: 0.1718323826789856
step: 240, loss: 0.05273817852139473
step: 250, loss: 0.08508866280317307
step: 260, loss: 0.024948136880993843
step: 270, loss: 0.02630997635424137
step: 280, loss: 0.0871419832110405
step: 290, loss: 0.1570618897676468
step: 300, loss: 0.11941925436258316
step: 310, loss: 0.10856889188289642
step: 320, loss: 0.04563376307487488
step: 330, loss: 0.09099269658327103
step: 340, loss: 0.11600308865308762
step: 350, loss: 0.061833128333091736
step: 360, loss: 0.08209402114152908
step: 370, loss: 0.17888905107975006
step: 380, loss: 0.10261429846286774
step: 390, loss: 0.018816906958818436
step: 400, loss: 0.0197516530752182
step: 410, loss: 0.08907116949558258
step: 420, loss: 0.09542643278837204
step: 430, loss: 0.14029823243618011
step: 440, loss: 0.17147605121135712
step: 450, loss: 0.06160837784409523
step: 460, loss: 0.047222282737493515
step: 470, loss: 0.322456032037735

```

step: 480, loss: 0.23830650746822357
step: 490, loss: 0.03472331911325455
step: 500, loss: 0.022319287061691284
step: 510, loss: 0.07134697586297989
step: 520, loss: 0.11214186251163483
step: 530, loss: 0.1164819598197937
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9217065166432256, f1=0.9165501165501165,
best_f1=0.9165501165501165
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07687433063983917
step: 10, loss: 0.10502351820468903
step: 20, loss: 0.07422841340303421
step: 30, loss: 0.022578120231628418
step: 40, loss: 0.019953802227973938
step: 50, loss: 0.1385396122932434
step: 60, loss: 0.2761226296424866
step: 70, loss: 0.08368732780218124
step: 80, loss: 0.018833061680197716
step: 90, loss: 0.016107695177197456
step: 100, loss: 0.22256837785243988
step: 110, loss: 0.04385669156908989
step: 120, loss: 0.03419583663344383
step: 130, loss: 0.022703742608428
step: 140, loss: 0.021353552117943764
step: 150, loss: 0.10656074434518814
step: 160, loss: 0.017078042030334473
step: 170, loss: 0.07624360173940659
step: 180, loss: 0.01623823493719101
step: 190, loss: 0.011626958847045898
step: 200, loss: 0.01364254578948021
step: 210, loss: 0.05080166086554527
step: 220, loss: 0.14354687929153442
step: 230, loss: 0.014943171292543411
step: 240, loss: 0.09938288480043411
step: 250, loss: 0.07233743369579315
step: 260, loss: 0.006323255132883787
step: 270, loss: 0.0841432437300682
step: 280, loss: 0.12813907861709595
step: 290, loss: 0.11437179148197174

```

```

step: 300, loss: 0.03221168741583824
step: 310, loss: 0.08125169575214386
step: 320, loss: 0.09849897772073746
step: 330, loss: 0.04136713966727257
step: 340, loss: 0.004770730156451464
step: 350, loss: 0.058354273438453674
step: 360, loss: 0.022163676097989082
step: 370, loss: 0.008445635437965393
step: 380, loss: 0.09392444789409637
step: 390, loss: 0.03230162709951401
step: 400, loss: 0.05574400722980499
step: 410, loss: 0.001438269391655922
step: 420, loss: 0.03195085749030113
step: 430, loss: 0.06116296723484993
step: 440, loss: 0.11864640563726425
step: 450, loss: 0.030057935044169426
step: 460, loss: 0.30538347363471985
step: 470, loss: 0.05958850681781769
step: 480, loss: 0.21956518292427063
step: 490, loss: 0.05587910860776901
step: 500, loss: 0.04030006751418114
step: 510, loss: 0.06314907222986221
step: 520, loss: 0.06559938937425613
step: 530, loss: 0.21530984342098236
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9273743016759776, f1=0.9230769230769231,
best_f1=0.9230769230769231
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.024289442226290703
step: 10, loss: 0.06749936193227768
step: 20, loss: 0.2085225135087967
step: 30, loss: 0.2417517751455307
step: 40, loss: 0.00372115895152092
step: 50, loss: 0.0059369951486587524
step: 60, loss: 0.031152117997407913
step: 70, loss: 0.13967235386371613
step: 80, loss: 0.015759453177452087
step: 90, loss: 0.048252105712890625
step: 100, loss: 0.02451162412762642
step: 110, loss: 0.0728701502084732

```

```
step: 120, loss: 0.04712523892521858
step: 130, loss: 0.00578541262075305
step: 140, loss: 0.02060014382004738
step: 150, loss: 0.013133158907294273
step: 160, loss: 0.030850673094391823
step: 170, loss: 0.002941520418971777
step: 180, loss: 0.08056320250034332
step: 190, loss: 0.00979346688836813
step: 200, loss: 0.07001195847988129
step: 210, loss: 0.07823894172906876
step: 220, loss: 0.015091250650584698
step: 230, loss: 0.013199621811509132
step: 240, loss: 0.006284950766712427
step: 250, loss: 0.011922095902264118
step: 260, loss: 0.0013617626391351223
step: 270, loss: 0.013633685186505318
step: 280, loss: 0.009493615478277206
step: 290, loss: 0.048931486904621124
step: 300, loss: 0.1960591971874237
step: 310, loss: 0.20155708491802216
step: 320, loss: 0.09466684609651566
step: 330, loss: 0.003642575815320015
step: 340, loss: 0.013334452174603939
step: 350, loss: 0.03270801529288292
step: 360, loss: 0.004726208746433258
step: 370, loss: 0.00640052231028676
step: 380, loss: 0.0048013306222856045
step: 390, loss: 0.05048491805791855
step: 400, loss: 0.014876020140945911
step: 410, loss: 0.0289083793759346
step: 420, loss: 0.07841674238443375
step: 430, loss: 0.04565054550766945
step: 440, loss: 0.06967251002788544
step: 450, loss: 0.13065296411514282
step: 460, loss: 0.048490434885025024
step: 470, loss: 0.05978856235742569
step: 480, loss: 0.004856497514992952
step: 490, loss: 0.013019482605159283
step: 500, loss: 0.12643516063690186
step: 510, loss: 0.0023711377289146185
step: 520, loss: 0.00804807711392641
step: 530, loss: 0.05156741291284561
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.933579335793358, f1=0.9208831646734131,
best_f1=0.9208831646734131
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0091166403144598
step: 10, loss: 0.038563475012779236
step: 20, loss: 0.003919986542314291
step: 30, loss: 0.0028192182071506977
step: 40, loss: 0.0165171530097723
step: 50, loss: 0.016706442460417747
step: 60, loss: 0.05124615132808685
step: 70, loss: 0.048107847571372986
step: 80, loss: 0.005661092232912779
step: 90, loss: 0.0665023997426033
step: 100, loss: 0.03212236985564232
step: 110, loss: 0.024337923154234886
step: 120, loss: 0.001720590516924858
step: 130, loss: 0.007892459630966187
step: 140, loss: 0.001045677694492042
step: 150, loss: 0.00068416737485677
step: 160, loss: 0.0033255338203161955
step: 170, loss: 0.004040207248181105
step: 180, loss: 0.002136934082955122
step: 190, loss: 0.006903655361384153
step: 200, loss: 0.0023197149857878685
step: 210, loss: 0.005066295620054007
step: 220, loss: 0.04963080212473869
step: 230, loss: 0.22846896946430206
step: 240, loss: 0.004606880713254213
step: 250, loss: 0.03060326538980007
step: 260, loss: 0.04265396296977997
step: 270, loss: 0.009065401740372181
step: 280, loss: 0.0018669397104531527
step: 290, loss: 0.005819864571094513
step: 300, loss: 0.0035562836565077305
step: 310, loss: 0.0016882626805454493
step: 320, loss: 0.006013599224388599
step: 330, loss: 0.10045702010393143
step: 340, loss: 0.02001032419502735
step: 350, loss: 0.0011716358130797744
step: 360, loss: 0.0026506937574595213
step: 370, loss: 0.1137162297964096
step: 380, loss: 0.0022404384799301624
step: 390, loss: 0.06404012441635132
step: 400, loss: 0.03346008062362671
step: 410, loss: 0.0018605293007567525

```

```

step: 420, loss: 0.007106696255505085
step: 430, loss: 0.010332132689654827
step: 440, loss: 0.03326326608657837
step: 450, loss: 0.015114140696823597
step: 460, loss: 0.018442675471305847
step: 470, loss: 0.02830885909497738
step: 480, loss: 0.0016039260663092136
step: 490, loss: 0.014384951442480087
step: 500, loss: 0.002447420498356223
step: 510, loss: 0.010482678189873695
step: 520, loss: 0.06192087382078171
step: 530, loss: 0.00284450501203537
/usr/local/lib/python3.7/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=0.9290681502086231, f1=0.9216674301420064,
best_f1=0.9208831646734131
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004896709229797125
step: 10, loss: 0.00951442588120699
step: 20, loss: 0.006615083198994398
step: 30, loss: 0.001008764491416514
step: 40, loss: 0.0012646537506952882
step: 50, loss: 0.006570738740265369
step: 60, loss: 0.00045704972581006587
step: 70, loss: 0.003468368900939822
step: 80, loss: 0.0015163661446422338
step: 90, loss: 0.016107117757201195
step: 100, loss: 0.027448272332549095
step: 110, loss: 0.0009739267406985164
step: 120, loss: 0.013680141419172287
step: 130, loss: 0.0034945455845445395
step: 140, loss: 0.002511161146685481
step: 150, loss: 0.018158266320824623
step: 160, loss: 0.144429549574852
step: 170, loss: 0.003354351967573166
step: 180, loss: 0.06967606395483017
step: 190, loss: 0.0012158756144344807
step: 200, loss: 0.044647522270679474
step: 210, loss: 0.0039430283941328526
step: 220, loss: 0.0011526246089488268
step: 230, loss: 0.0034301732666790485
step: 240, loss: 0.003350981045514345
step: 250, loss: 0.0010005328804254532
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 260, loss: 0.0917288139462471

```

```

step: 270, loss: 0.010085226967930794
step: 280, loss: 0.09088826924562454
step: 290, loss: 0.0009824453154578805
step: 300, loss: 0.14754217863082886
step: 310, loss: 0.0027355642523616552
step: 320, loss: 0.002638606121763587
step: 330, loss: 0.024600930511951447
step: 340, loss: 0.001069317921064794
step: 350, loss: 0.0764666199684143
step: 360, loss: 0.009397258050739765
step: 370, loss: 0.008513634093105793
step: 380, loss: 0.02454100176692009
step: 390, loss: 0.017481882125139236
step: 400, loss: 0.01814166083931923
step: 410, loss: 0.002937944373115897
step: 420, loss: 0.0013585189590230584
step: 430, loss: 0.002301124855875969
step: 440, loss: 0.009869069792330265
step: 450, loss: 0.05372069403529167
step: 460, loss: 0.005935617722570896
step: 470, loss: 0.0008052214398048818
step: 480, loss: 0.0013534664176404476
step: 490, loss: 0.010462385602295399
step: 500, loss: 0.005743204616010189
step: 510, loss: 0.0034235736820846796
step: 520, loss: 0.0009356985101476312
step: 530, loss: 0.023394932970404625
/usr/local/lib/python3.7/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=0.9298724954462659, f1=0.9252378794743996,
best_f1=0.9208831646734131
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008314087754115462
step: 10, loss: 0.0020436819177120924
step: 20, loss: 0.0009666242985986173
step: 30, loss: 0.0025899256579577923
step: 40, loss: 0.001087798154912889
step: 50, loss: 0.07261820882558823
step: 60, loss: 0.0026868050917983055
step: 70, loss: 0.01956626959145069
step: 80, loss: 0.00999732967466116
step: 90, loss: 0.014985630288720131
step: 100, loss: 0.13428351283073425
step: 110, loss: 0.03422041982412338
step: 120, loss: 0.00046478997683152556

```



```

step: 130, loss: 0.00025606248527765274
step: 140, loss: 0.0002697053423617035
step: 150, loss: 0.014136663638055325
step: 160, loss: 0.003234586212784052
step: 170, loss: 0.00026156066451221704
step: 180, loss: 0.00012431635695975274
step: 190, loss: 0.00016676368250045925
step: 200, loss: 0.00013926380779594183
step: 210, loss: 0.000845389673486352
step: 220, loss: 0.00014120583364274353
step: 230, loss: 0.0003171396383550018
step: 240, loss: 0.0002470730687491596
step: 250, loss: 0.03784160688519478
step: 260, loss: 0.002486184472218156
step: 270, loss: 0.0015736715868115425
step: 280, loss: 0.002699376316741109
step: 290, loss: 0.005589098669588566
step: 300, loss: 0.0006976763834245503
step: 310, loss: 0.0014572525396943092
step: 320, loss: 0.06543704867362976
step: 330, loss: 0.028781799599528313
step: 340, loss: 0.0016191666945815086
step: 350, loss: 0.10531747341156006
step: 360, loss: 0.0028444258496165276
step: 370, loss: 0.0028773292433470488
step: 380, loss: 0.003435005433857441
step: 390, loss: 0.032018937170505524
step: 400, loss: 0.0001877641916507855
step: 410, loss: 0.00031566820689477026
step: 420, loss: 0.005300832912325859
step: 430, loss: 0.0013188477605581284
step: 440, loss: 0.0024532293900847435
step: 450, loss: 0.003925587050616741
step: 460, loss: 0.13042820990085602
step: 470, loss: 0.09599688649177551
step: 480, loss: 0.05161314085125923
step: 490, loss: 0.0005359124043025076
step: 500, loss: 0.009060938842594624
step: 510, loss: 0.0004037751059513539
step: 520, loss: 0.0007015072624199092
step: 530, loss: 0.003499801503494382
/usr/local/lib/python3.7/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=0.9285042333019755, f1=0.9227871939736347,
best_f1=0.9208831646734131
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04085627198219299
step: 10, loss: 0.007024372462183237
step: 20, loss: 0.0034526530653238297
step: 30, loss: 0.0024448733311146498
step: 40, loss: 0.0001738544669933617
step: 50, loss: 0.0004939974169246852
step: 60, loss: 0.08956244587898254
step: 70, loss: 0.005745938513427973
step: 80, loss: 0.002367559354752302
step: 90, loss: 0.00014099563122726977
step: 100, loss: 0.00014727101370226592
step: 110, loss: 0.05966987460851669
step: 120, loss: 0.0003671289305202663
step: 130, loss: 0.00013841092004440725
step: 140, loss: 0.0005168973584659398
step: 150, loss: 0.00012806191807612777
step: 160, loss: 0.0005687209195457399
step: 170, loss: 0.001300192903727293
step: 180, loss: 0.00017511771875433624
step: 190, loss: 0.0002638738078530878
step: 200, loss: 0.00016048448742367327
step: 210, loss: 0.0016522642690688372
step: 220, loss: 9.313004557043314e-05
step: 230, loss: 0.004005619790405035
step: 240, loss: 0.005120898596942425
step: 250, loss: 0.01784096658229828
step: 260, loss: 0.0017177439294755459
step: 270, loss: 6.930073141120374e-05
step: 280, loss: 0.04087941348552704
step: 290, loss: 0.002765857381746173
step: 300, loss: 8.151614747475833e-05
step: 310, loss: 0.00014229805674403906
step: 320, loss: 0.01420519221574068
step: 330, loss: 9.397754911333323e-05
step: 340, loss: 0.022994551807641983
step: 350, loss: 0.0008577140979468822
step: 360, loss: 0.0045182243920862675
step: 370, loss: 0.0009325661230832338
step: 380, loss: 0.002154394518584013
step: 390, loss: 6.257397762965411e-05
step: 400, loss: 0.0005317408940754831
step: 410, loss: 9.85184742603451e-05
step: 420, loss: 0.00012470237561501563
step: 430, loss: 0.00011851680028485134
step: 440, loss: 0.0002123218437191099
step: 450, loss: 0.0008107466856017709
step: 460, loss: 0.0021831714548170567
```

```

step: 470, loss: 0.004808816127479076
step: 480, loss: 0.00023104646243155003
step: 490, loss: 0.0005028999876230955
step: 500, loss: 0.0003965092182625085
step: 510, loss: 0.0009068168001249433
step: 520, loss: 0.0003043391043320298
step: 530, loss: 0.0005716775194741786
/usr/local/lib/python3.7/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=0.9330855018587362, f1=0.9231481481481482,
best_f1=0.9208831646734131
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0016182672698050737
step: 10, loss: 0.007572160102427006
step: 20, loss: 0.016423163935542107
step: 30, loss: 0.00027816276997327805
step: 40, loss: 0.0002996924158651382
step: 50, loss: 0.0003541111364029348
step: 60, loss: 0.012414194643497467
step: 70, loss: 0.06887491047382355
step: 80, loss: 0.00011006152635673061
step: 90, loss: 0.0006554066785611212
step: 100, loss: 0.006719219498336315
step: 110, loss: 0.000606912886723876
step: 120, loss: 0.0001668606564635411
step: 130, loss: 0.0004720492579508573
step: 140, loss: 4.995462950319052e-05
step: 150, loss: 0.0015592476120218635
step: 160, loss: 9.909793880069628e-05
step: 170, loss: 0.0011637448333203793
step: 180, loss: 0.0006519323214888573
step: 190, loss: 0.00020447249698918313
step: 200, loss: 0.006168627645820379
step: 210, loss: 0.05726541206240654
step: 220, loss: 0.001901406329125166
step: 230, loss: 0.0003809587506111711
step: 240, loss: 0.00028467964148148894
step: 250, loss: 0.0011331699788570404
step: 260, loss: 0.023396048694849014
step: 270, loss: 8.159642311511561e-05
step: 280, loss: 0.00043676013592630625
step: 290, loss: 0.039627522230148315
step: 300, loss: 0.0005249275709502399
step: 310, loss: 0.01108712237328291
step: 320, loss: 0.0011325681116431952

```

```

step: 330, loss: 0.022977614775300026
step: 340, loss: 0.0001535000919830054
step: 350, loss: 0.007567386608570814
step: 360, loss: 0.0018404475413262844
step: 370, loss: 0.005265503656119108
step: 380, loss: 0.015602118335664272
step: 390, loss: 0.00029830296989530325
step: 400, loss: 0.04772434011101723
step: 410, loss: 0.0001547202409710735
step: 420, loss: 0.00043867729254998267
step: 430, loss: 0.002064415952190757
step: 440, loss: 0.0002699344768188894
step: 450, loss: 0.0004506891709752381
step: 460, loss: 0.0020215720869600773
step: 470, loss: 0.0007360352319665253
step: 480, loss: 0.0003613409644458443
step: 490, loss: 0.00012591791164595634
step: 500, loss: 0.0022143349051475525
step: 510, loss: 7.547537097707391e-05
step: 520, loss: 0.0003564766375347972
step: 530, loss: 0.0026932964101433754
/usr/local/lib/python3.7/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=0.9310344827586207, f1=0.9239981575310916,
best_f1=0.9208831646734131
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004717460833489895
step: 10, loss: 0.017415419220924377
step: 20, loss: 0.00040613237069919705
step: 30, loss: 9.544157364871353e-05
step: 40, loss: 0.0004049840208608657
step: 50, loss: 0.025899698957800865
step: 60, loss: 0.0001371231919620186
step: 70, loss: 0.059418875724077225
step: 80, loss: 0.000553426390979439
step: 90, loss: 6.348378519760445e-05
step: 100, loss: 0.00013854372082278132
step: 110, loss: 0.12289778888225555
step: 120, loss: 0.0002609398798085749
step: 130, loss: 0.0007565804990008473
step: 140, loss: 0.00017476899665780365
step: 150, loss: 0.022960208356380463
step: 160, loss: 0.00022672078921459615
step: 170, loss: 0.00013182306429371238
step: 180, loss: 0.00021583375928457826

```

```

step: 190, loss: 0.000144641162478365
step: 200, loss: 0.0005949243786744773
step: 210, loss: 4.91229111503344e-05
step: 220, loss: 0.004687547218054533
step: 230, loss: 0.00023581003188155591
step: 240, loss: 5.313778456184082e-05
step: 250, loss: 0.0001882987271528691
step: 260, loss: 0.0013401485048234463
step: 270, loss: 0.002109413966536522
step: 280, loss: 0.0001554749032948166
step: 290, loss: 0.0010483040241524577
step: 300, loss: 0.00032052036840468645
step: 310, loss: 0.0009069616789929569
step: 320, loss: 0.0012988096568733454
step: 330, loss: 0.019403601065278053
step: 340, loss: 0.00033446887391619384
step: 350, loss: 0.00010536891204537824
step: 360, loss: 0.003686717711389065
step: 370, loss: 0.00047853298019617796
step: 380, loss: 7.874107541283593e-05
step: 390, loss: 0.0003433184465393424
step: 400, loss: 0.0021999769378453493
step: 410, loss: 0.00017865246627479792
step: 420, loss: 8.856668137013912e-05
step: 430, loss: 3.8226724427659065e-05
step: 440, loss: 0.00024549299268983305
step: 450, loss: 4.771781459567137e-05
step: 460, loss: 0.0016590430168434978
step: 470, loss: 0.00010874681174755096
step: 480, loss: 0.007587088271975517
step: 490, loss: 7.029034895822406e-05
step: 500, loss: 0.006028566975146532
step: 510, loss: 0.03323163092136383
step: 520, loss: 4.2688472603913397e-05
step: 530, loss: 6.074114571674727e-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=0.92616226071103, f1=0.919854280510018,
best_f1=0.9208831646734131
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000386866187909618
step: 10, loss: 0.0016124679241329432
step: 20, loss: 5.11233156430535e-05
step: 30, loss: 8.786933176452294e-05
step: 40, loss: 0.00014163856394588947

```

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

step: 50, loss: 3.251332600484602e-05
step: 60, loss: 3.53519462805707e-05
step: 70, loss: 8.286424417747185e-05
step: 80, loss: 0.00470053730532527
step: 90, loss: 4.0931619878392667e-05
step: 100, loss: 0.000972378475125879
step: 110, loss: 0.001789859146811068
step: 120, loss: 3.4442935429979116e-05
step: 130, loss: 2.8989543352508917e-05
step: 140, loss: 0.0002846091229002923
step: 150, loss: 9.018702985486016e-05
step: 160, loss: 0.0004071812436450273
step: 170, loss: 0.031924135982990265
step: 180, loss: 0.007589933928102255
step: 190, loss: 4.6656808990519494e-05
step: 200, loss: 0.0003099109744653106
step: 210, loss: 0.0014692104887217283
step: 220, loss: 0.00399838387966156
step: 230, loss: 4.899127088719979e-05
step: 240, loss: 0.05572966858744621
step: 250, loss: 0.00016667736053932458
step: 260, loss: 0.0195162370800972
step: 270, loss: 2.9030175937805325e-05
step: 280, loss: 3.8666639738949016e-05
step: 290, loss: 0.0002776970504783094
step: 300, loss: 7.994897896423936e-05
step: 310, loss: 0.0001221008860738948
step: 320, loss: 5.4870735766598955e-05
step: 330, loss: 0.07103624939918518
step: 340, loss: 0.00012513087131083012
step: 350, loss: 0.0019652375485748053
step: 360, loss: 5.687962402589619e-05
step: 370, loss: 0.0006999171455390751
step: 380, loss: 0.00040288257878273726
step: 390, loss: 4.536429696599953e-05
step: 400, loss: 0.003626021323725581
step: 410, loss: 0.0031925695948302746
step: 420, loss: 0.004136353265494108
step: 430, loss: 3.580970223993063e-05
step: 440, loss: 8.491826156387106e-05
step: 450, loss: 0.00016739025886636227
step: 460, loss: 0.00042513193329796195
step: 470, loss: 3.7016801798017696e-05
step: 480, loss: 8.665181667311117e-05
step: 490, loss: 0.055800702422857285
step: 500, loss: 0.00036058228579349816
step: 510, loss: 0.001624943921342492

```

step: 520, loss: 0.002634721575304866
step: 530, loss: 0.00016740408318582922
/usr/local/lib/python3.7/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=0.9268065268065269, f1=0.9217954650624711,
best_f1=0.9208831646734131
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003212386800441891
step: 10, loss: 0.0009690124425105751
step: 20, loss: 0.0001205865919473581
step: 30, loss: 0.00016107133706100285
step: 40, loss: 0.0006508245132863522
step: 50, loss: 0.0006471130764111876
step: 60, loss: 0.0003036189591512084
step: 70, loss: 8.556011744076386e-05
step: 80, loss: 0.0005568161723203957
step: 90, loss: 7.099861977621913e-05
step: 100, loss: 0.00018844546866603196
step: 110, loss: 7.279749115696177e-05
step: 120, loss: 9.944835619535297e-05
step: 130, loss: 0.0002256790321553126
step: 140, loss: 0.006692708004266024
step: 150, loss: 0.00028575301985256374
step: 160, loss: 0.00025135991745628417
step: 170, loss: 0.0003575061564333737
step: 180, loss: 3.8808357203379273e-05
step: 190, loss: 5.55701190023683e-05
step: 200, loss: 0.00026828848058357835
step: 210, loss: 7.784971967339516e-05
step: 220, loss: 4.113658724236302e-05
step: 230, loss: 3.6674082366516814e-05
step: 240, loss: 4.731459921458736e-05
step: 250, loss: 0.0003172008437104523
step: 260, loss: 0.0002604158944450319
step: 270, loss: 0.00019621413957793266
step: 280, loss: 6.398481491487473e-05
step: 290, loss: 0.00015732347674202174
step: 300, loss: 0.09635745733976364
step: 310, loss: 0.0009048313368111849
step: 320, loss: 0.009091329760849476
step: 330, loss: 0.008397500030696392
step: 340, loss: 0.000539364933501929
step: 350, loss: 0.0019079938065260649
step: 360, loss: 0.0001522358797956258
step: 370, loss: 0.00015378245734609663

```

```
step: 380, loss: 6.768936873413622e-05
step: 390, loss: 0.00014961422129999846
step: 400, loss: 3.264323095208965e-05
step: 410, loss: 3.155523882014677e-05
step: 420, loss: 0.0003449745418038219
step: 430, loss: 0.0004003611975349486
step: 440, loss: 4.5534761738963425e-05
step: 450, loss: 3.1943411158863455e-05
step: 460, loss: 0.00036037852987647057
step: 470, loss: 0.0007679421687498689
step: 480, loss: 3.326211299281567e-05
step: 490, loss: 0.22557294368743896
step: 500, loss: 0.00012405570305418223
step: 510, loss: 9.152325219474733e-05
step: 520, loss: 0.0002122859877999872
step: 530, loss: 6.94388581905514e-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=0.923581809657759, f1=0.9182156133828996,
best_f1=0.9208831646734131
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 6.406367174349725e-05
step: 10, loss: 0.003838162636384368
step: 20, loss: 0.00014897424262017012
step: 30, loss: 0.0002767493133433163
step: 40, loss: 0.0015392140485346317
step: 50, loss: 6.878867134219036e-05
step: 60, loss: 7.904623635113239e-05
step: 70, loss: 0.00028645203565247357
step: 80, loss: 8.805270044831559e-05
step: 90, loss: 4.1341703763464466e-05
step: 100, loss: 0.0011765643721446395
step: 110, loss: 6.378196849254891e-05
step: 120, loss: 0.00816788524389267
step: 130, loss: 0.000141557480674237
step: 140, loss: 5.2906794735463336e-05
step: 150, loss: 0.00022010286920703948
step: 160, loss: 9.008000051835552e-05
step: 170, loss: 3.098603701801039e-05
step: 180, loss: 0.0002647563233040273
step: 190, loss: 0.012389901094138622
step: 200, loss: 4.761418676935136e-05
step: 210, loss: 7.902743527665734e-05
step: 220, loss: 3.4908720408566296e-05
step: 230, loss: 0.000330794311594218
```



```

step: 240, loss: 4.222102870699018e-05
step: 250, loss: 7.209562318166718e-05
step: 260, loss: 0.000178576898179017
step: 270, loss: 4.391013862914406e-05
step: 280, loss: 6.800160917919129e-05
step: 290, loss: 0.000830710690934211
step: 300, loss: 0.00012756221985910088
step: 310, loss: 3.878259303746745e-05
step: 320, loss: 5.0566137360874563e-05
step: 330, loss: 3.777253004955128e-05
step: 340, loss: 9.566832886775956e-05
step: 350, loss: 0.0016255829250440001
step: 360, loss: 0.00021390430629253387
step: 370, loss: 4.18991367041599e-05
step: 380, loss: 6.979372119531035e-05
step: 390, loss: 4.027226896141656e-05
step: 400, loss: 8.040686225285754e-05
step: 410, loss: 5.8313515182817355e-05
step: 420, loss: 0.00017026023124344647
step: 430, loss: 6.453431706177071e-05
step: 440, loss: 0.00011019988596672192
step: 450, loss: 0.0004246324533596635
step: 460, loss: 0.0016275016823783517
step: 470, loss: 0.00040092680137604475
step: 480, loss: 0.009283039718866348
step: 490, loss: 3.974874925916083e-05
step: 500, loss: 0.008915478363633156
step: 510, loss: 0.0003201024665031582
step: 520, loss: 0.0003513716801535338
step: 530, loss: 0.0026081118267029524
/usr/local/lib/python3.7/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=0.926852288815479, f1=0.9196806012212306,
best_f1=0.9208831646734131
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.734778147190809e-05
step: 10, loss: 2.8572005248861387e-05
step: 20, loss: 0.001224043662659824
step: 30, loss: 0.0010841533076018095
step: 40, loss: 7.570772868348286e-05
step: 50, loss: 0.00010306061449227855
step: 60, loss: 2.983506055898033e-05
step: 70, loss: 2.389724977547303e-05
step: 80, loss: 2.424372905807104e-05
step: 90, loss: 0.000699387863278389

```

```
step: 100, loss: 2.5193523470079526e-05
step: 110, loss: 0.00095345702720806
step: 120, loss: 4.2675139411585405e-05
step: 130, loss: 6.603432120755315e-05
step: 140, loss: 5.5778564274078235e-05
step: 150, loss: 6.201510404935107e-05
step: 160, loss: 2.3219357899506576e-05
step: 170, loss: 4.704440289060585e-05
step: 180, loss: 3.0039782359381206e-05
step: 190, loss: 5.058505848865025e-05
step: 200, loss: 0.004753924906253815
step: 210, loss: 0.0010696372482925653
step: 220, loss: 4.234054358676076e-05
step: 230, loss: 0.000293257791781798
step: 240, loss: 0.0003318304370623082
step: 250, loss: 3.878592178807594e-05
step: 260, loss: 2.3830205464037135e-05
step: 270, loss: 6.0809648857684806e-05
step: 280, loss: 3.657585693872534e-05
step: 290, loss: 0.00013042001228313893
step: 300, loss: 0.0001802261103875935
step: 310, loss: 2.2004867787472904e-05
step: 320, loss: 3.0293125746538863e-05
step: 330, loss: 3.4383181628072634e-05
step: 340, loss: 2.936191958724521e-05
step: 350, loss: 0.0012937032151967287
step: 360, loss: 4.462023571250029e-05
step: 370, loss: 2.499196671124082e-05
step: 380, loss: 6.581445632036775e-05
step: 390, loss: 4.625451038009487e-05
step: 400, loss: 0.0010347269708290696
step: 410, loss: 2.7953765311394818e-05
step: 420, loss: 8.823168172966689e-05
step: 430, loss: 5.028849409427494e-05
step: 440, loss: 3.291896791779436e-05
step: 450, loss: 3.9637714507989585e-05
step: 460, loss: 2.3502301701228134e-05
step: 470, loss: 5.558213160838932e-05
step: 480, loss: 2.610989940876607e-05
step: 490, loss: 8.308325777761638e-05
step: 500, loss: 3.310203828732483e-05
step: 510, loss: 0.0007561969687230885
step: 520, loss: 3.3357704523950815e-05
step: 530, loss: 3.181617648806423e-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=0.9279999999999999, f1=0.9233662435354959,
```

```
best_f1=0.9208831646734131
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.655430191429332e-05
step: 10, loss: 6.914994444656074e-05
step: 20, loss: 6.946636858629063e-05
step: 30, loss: 5.4572268709307536e-05
step: 40, loss: 3.695349732879549e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 50, loss: 0.16825272142887115
step: 60, loss: 3.693110556923784e-05
step: 70, loss: 2.7443133149063215e-05
step: 80, loss: 3.876331902574748e-05
step: 90, loss: 0.00015568267554044724
step: 100, loss: 8.448095468338579e-05
step: 110, loss: 0.0003914213157258928
step: 120, loss: 1.8208986148238182e-05
step: 130, loss: 3.383124931133352e-05
step: 140, loss: 0.00015639139746781439
step: 150, loss: 3.413006925256923e-05
step: 160, loss: 0.0002513242943678051
step: 170, loss: 9.109629172598943e-05
step: 180, loss: 0.0007826110231690109
step: 190, loss: 8.099564001895487e-05
step: 200, loss: 4.155260103289038e-05
step: 210, loss: 6.159839540487155e-05
step: 220, loss: 4.9336147640133277e-05
step: 230, loss: 5.382444942370057e-05
step: 240, loss: 4.799925227416679e-05
step: 250, loss: 7.581148383906111e-05
step: 260, loss: 2.3249209334608167e-05
step: 270, loss: 0.00044676000834442675
step: 280, loss: 3.883436511387117e-05
step: 290, loss: 0.0002704451617319137
step: 300, loss: 0.00010787177598103881
step: 310, loss: 2.6329842512495816e-05
step: 320, loss: 6.713398761348799e-05
step: 330, loss: 2.5703999199322425e-05
step: 340, loss: 5.224556298344396e-05
step: 350, loss: 0.0003926404460798949
step: 360, loss: 0.004568048287183046
step: 370, loss: 0.00021314564219210297
step: 380, loss: 0.0002161542361136526
step: 390, loss: 0.0006533988635055721
step: 400, loss: 3.6674155126092955e-05
step: 410, loss: 3.8480706280097365e-05
step: 420, loss: 3.8156209484441206e-05
```

```
step: 430, loss: 4.661044295062311e-05
step: 440, loss: 2.0548433894873597e-05
step: 450, loss: 3.9732443838147447e-05
step: 460, loss: 0.0003690429439302534
step: 470, loss: 3.2285555789712816e-05
step: 480, loss: 2.8959741030121222e-05
step: 490, loss: 6.136802403489128e-05
step: 500, loss: 5.6886801758082584e-05
step: 510, loss: 6.257990753510967e-05
step: 520, loss: 3.425309478188865e-05
step: 530, loss: 0.00010075124009745196
/usr/local/lib/python3.7/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=0.9294062646096307, f1=0.9278734295020938,
best_f1=0.9208831646734131
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002760939532890916
step: 10, loss: 0.00015397454262711108
step: 20, loss: 0.0012843060540035367
step: 30, loss: 0.0003311007167212665
step: 40, loss: 2.7625814254861325e-05
step: 50, loss: 4.910357529297471e-05
step: 60, loss: 2.2682923372485675e-05
step: 70, loss: 2.8136224500485696e-05
step: 80, loss: 9.409264021087438e-05
step: 90, loss: 3.386859316378832e-05
step: 100, loss: 3.1563049560645595e-05
step: 110, loss: 0.00014093298523221165
step: 120, loss: 0.00022805147455073893
step: 130, loss: 0.00011494204227346927
step: 140, loss: 2.273509198857937e-05
step: 150, loss: 0.0006393579533323646
step: 160, loss: 4.338663711678237e-05
step: 170, loss: 0.00032519508386030793
step: 180, loss: 4.024852751172148e-05
step: 190, loss: 4.645589069696143e-05
step: 200, loss: 4.152621477260254e-05
step: 210, loss: 3.156654929625802e-05
step: 220, loss: 2.953621333290357e-05
step: 230, loss: 2.5514014851069078e-05
step: 240, loss: 2.9902208552812226e-05
step: 250, loss: 8.520435949321836e-05
step: 260, loss: 0.0003806464665103704
step: 270, loss: 4.156802606303245e-05
step: 280, loss: 0.00017278827726840973
```

```

step: 290, loss: 0.00033030862687155604
step: 300, loss: 0.000240965629927814
step: 310, loss: 3.512437251629308e-05
step: 320, loss: 4.67442259832751e-05
step: 330, loss: 0.0001124337431974709
step: 340, loss: 0.039115529507398605
step: 350, loss: 4.2471543565625325e-05
step: 360, loss: 3.3291969884885475e-05
step: 370, loss: 2.0235504052834585e-05
step: 380, loss: 3.4300970582989976e-05
step: 390, loss: 3.0959814466768876e-05
step: 400, loss: 3.2350981200579554e-05
step: 410, loss: 0.00041523517575114965
step: 420, loss: 4.1452611185377464e-05
step: 430, loss: 2.563696398283355e-05
step: 440, loss: 7.024371734587476e-05
step: 450, loss: 0.0001346915669273585
step: 460, loss: 2.4631073756609112e-05
step: 470, loss: 2.5566014301148243e-05
step: 480, loss: 0.00036939253914169967
step: 490, loss: 9.47310181800276e-05
step: 500, loss: 3.11091062030755e-05
step: 510, loss: 0.00019870258984155953
step: 520, loss: 3.340562034281902e-05
step: 530, loss: 3.581331475288607e-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=0.9280742459396751, f1=0.925207756232687,
best_f1=0.9208831646734131
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

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

```

```

keep_batchnorm_fp32      : True
master_weights           : True
loss_scale                : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
5742it [00:22, 257.39it/s]
load_f1 = 0.9345794392523366
real_f1 = 0.930776426566885
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:17, 255.97it/s]

```

2.1.4 Amazon-Google - Running the matcher

```

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

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

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

```

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

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

Defaults for this optimization level are:

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

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

After processing overrides, optimization options are:

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

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

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

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

An input tensor was not cuda.

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

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

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

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

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

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

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

step: 0, loss: 0.8303253054618835

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

step: 10, loss: 0.06305339187383652

step: 20, loss: 0.3771224021911621

step: 30, loss: 0.37202370166778564

step: 40, loss: 0.5038699507713318

step: 50, loss: 0.2901965081691742

```

step: 60, loss: 0.3525298237800598
step: 70, loss: 0.23978814482688904
step: 80, loss: 0.30521804094314575
step: 90, loss: 0.3847125470638275
step: 100, loss: 0.15434573590755463
step: 110, loss: 0.2819751501083374
step: 120, loss: 0.23561802506446838
step: 130, loss: 0.20704877376556396
step: 140, loss: 0.20914363861083984
step: 150, loss: 0.22848519682884216
step: 160, loss: 0.19981329143047333
step: 170, loss: 0.15909574925899506
step: 180, loss: 0.20157036185264587
step: 190, loss: 0.2738252580165863
step: 200, loss: 0.13552837073802948
step: 210, loss: 0.3541596531867981
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5563139931740614, f1=0.5779334500875657,
best_f1=0.5779334500875657
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07482714205980301
step: 10, loss: 0.06680580973625183
step: 20, loss: 0.21510618925094604
step: 30, loss: 0.12450120598077774
step: 40, loss: 0.12297911942005157
step: 50, loss: 0.15499192476272583
step: 60, loss: 0.03972739353775978
step: 70, loss: 0.17903298139572144
step: 80, loss: 0.12997831404209137
step: 90, loss: 0.122475765645504
step: 100, loss: 0.07089361548423767
step: 110, loss: 0.09188876301050186
step: 120, loss: 0.16641715168952942
step: 130, loss: 0.20786073803901672
step: 140, loss: 0.26196521520614624
step: 150, loss: 0.16421182453632355
step: 160, loss: 0.08062653243541718
step: 170, loss: 0.16911697387695312
step: 180, loss: 0.299400269985199
step: 190, loss: 0.18152739107608795

```



```

step: 200, loss: 0.1523386687040329
step: 210, loss: 0.18688204884529114
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6119096509240245, f1=0.631578947368421,
best_f1=0.631578947368421
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1891404688358307
step: 10, loss: 0.21007536351680756
step: 20, loss: 0.21825675666332245
step: 30, loss: 0.07673252373933792
step: 40, loss: 0.07194952666759491
step: 50, loss: 0.15083368122577667
step: 60, loss: 0.17180083692073822
step: 70, loss: 0.12689334154129028
step: 80, loss: 0.078921839594841
step: 90, loss: 0.050469838082790375
step: 100, loss: 0.0070401765406131744
step: 110, loss: 0.10000171512365341
step: 120, loss: 0.13200969994068146
step: 130, loss: 0.10718457400798798
step: 140, loss: 0.12508068978786469
step: 150, loss: 0.07926031202077866
step: 160, loss: 0.03584516420960426
step: 170, loss: 0.2859947681427002
step: 180, loss: 0.2248222529888153
step: 190, loss: 0.3677721917629242
step: 200, loss: 0.11650368571281433
step: 210, loss: 0.12147116661071777
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6139359698681732, f1=0.6041275797373358,
best_f1=0.6041275797373358
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.07596860080957413
step: 10, loss: 0.026299582794308662
step: 20, loss: 0.07727953791618347
step: 30, loss: 0.03552793338894844
step: 40, loss: 0.058589473366737366
step: 50, loss: 0.0904352068901062
step: 60, loss: 0.07255468517541885
step: 70, loss: 0.09807196259498596
step: 80, loss: 0.14329446852207184
step: 90, loss: 0.05072830617427826
step: 100, loss: 0.18064796924591064
step: 110, loss: 0.15743090212345123
step: 120, loss: 0.1584567129611969
step: 130, loss: 0.08582037687301636
step: 140, loss: 0.11774907261133194
step: 150, loss: 0.0994066372513771
step: 160, loss: 0.11786617338657379
step: 170, loss: 0.09189300239086151
step: 180, loss: 0.1587478369474411
step: 190, loss: 0.04962523281574249
step: 200, loss: 0.03667839989066124
step: 210, loss: 0.012759501114487648
/usr/local/lib/python3.7/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=0.5979381443298969, f1=0.5878378378378379,
best_f1=0.6041275797373358
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01741217076778412
step: 10, loss: 0.05794664844870567
step: 20, loss: 0.00800749845802784
step: 30, loss: 0.006788373924791813
step: 40, loss: 0.09576911479234695
step: 50, loss: 0.26354333758354187
step: 60, loss: 0.04341476410627365
step: 70, loss: 0.10693979263305664
step: 80, loss: 0.05597645416855812
step: 90, loss: 0.06884249299764633
step: 100, loss: 0.029205238446593285
step: 110, loss: 0.0034309360198676586
step: 120, loss: 0.0069397566840052605
step: 130, loss: 0.11494481563568115
step: 140, loss: 0.007203251123428345
step: 150, loss: 0.046856604516506195
step: 160, loss: 0.007850059308111668
step: 170, loss: 0.15449422597885132

```

```

step: 180, loss: 0.09611737728118896
step: 190, loss: 0.1422073096036911
step: 200, loss: 0.04579474776983261
step: 210, loss: 0.061220213770866394
/usr/local/lib/python3.7/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=0.5975609756097561, f1=0.5953307392996109,
best_f1=0.6041275797373358
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.054707758128643036
step: 10, loss: 0.015075970441102982
step: 20, loss: 0.01219249702990055
step: 30, loss: 0.1290809065103531
step: 40, loss: 0.0189052764326334
step: 50, loss: 0.039288293570280075
step: 60, loss: 0.20903900265693665
step: 70, loss: 0.0012974378187209368
step: 80, loss: 0.049462683498859406
step: 90, loss: 0.09697341173887253
step: 100, loss: 0.06617595255374908
step: 110, loss: 0.05182591825723648
step: 120, loss: 0.014646871946752071
step: 130, loss: 0.035377394407987595
step: 140, loss: 0.03350745886564255
step: 150, loss: 0.013559226877987385
step: 160, loss: 0.23851075768470764
step: 170, loss: 0.03898914158344269
step: 180, loss: 0.0034987207036465406
step: 190, loss: 0.0892328992486
step: 200, loss: 0.011975754052400589
step: 210, loss: 0.00854362454265356
/usr/local/lib/python3.7/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=0.6070175438596491, f1=0.6079447322970639,
best_f1=0.6041275797373358
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02886556275188923
step: 10, loss: 0.024393761530518532
step: 20, loss: 0.01565391570329666
step: 30, loss: 0.005484489724040031
step: 40, loss: 0.031872957944869995
step: 50, loss: 0.007440634537488222

```

```

step: 60, loss: 0.096466064453125
step: 70, loss: 0.006353581789880991
step: 80, loss: 0.005418572574853897
step: 90, loss: 0.029547380283474922
step: 100, loss: 0.028473228216171265
step: 110, loss: 0.006501017604023218
step: 120, loss: 0.020928381010890007
step: 130, loss: 0.041892506182193756
step: 140, loss: 0.0020089568570256233
step: 150, loss: 0.02028479613363743
step: 160, loss: 0.017645547166466713
step: 170, loss: 0.0146131981164217
step: 180, loss: 0.00032958644442260265
step: 190, loss: 0.0915064811706543
step: 200, loss: 0.01980261132121086
step: 210, loss: 0.007433632388710976
/usr/local/lib/python3.7/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=0.5960502692998205, f1=0.5957446808510638,
best_f1=0.6041275797373358
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.015347628854215145
step: 10, loss: 0.007448303978890181
step: 20, loss: 0.00837224256247282
step: 30, loss: 0.004588397219777107
step: 40, loss: 0.1645582616329193
step: 50, loss: 0.007990405894815922
step: 60, loss: 0.05495046079158783
step: 70, loss: 0.0006628716946579516
step: 80, loss: 0.025438372045755386
step: 90, loss: 0.10041870921850204
step: 100, loss: 0.0009479069849476218
step: 110, loss: 0.001896034344099462
step: 120, loss: 0.009112306870520115
step: 130, loss: 0.0014865786070004106
step: 140, loss: 0.0013537819031625986
step: 150, loss: 0.002689911052584648
step: 160, loss: 0.002791796112433076
step: 170, loss: 0.011440434493124485
step: 180, loss: 0.07724146544933319
step: 190, loss: 0.14477097988128662
step: 200, loss: 0.12413299083709717
step: 210, loss: 0.027280766516923904
/usr/local/lib/python3.7/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=0.5798816568047337, f1=0.6297029702970298,
best_f1=0.6041275797373358
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010304267518222332
step: 10, loss: 0.011022966355085373
step: 20, loss: 0.00705381017178297
step: 30, loss: 0.2366849184036255
step: 40, loss: 0.019255423918366432
step: 50, loss: 0.0014535585651174188
step: 60, loss: 0.0027300063520669937
step: 70, loss: 0.06376606225967407
step: 80, loss: 0.024986352771520615
step: 90, loss: 0.014187097549438477
step: 100, loss: 0.0058588809333741665
step: 110, loss: 0.019327284768223763
step: 120, loss: 0.0015276779886335135
step: 130, loss: 0.006904693320393562
step: 140, loss: 0.007197620812803507
step: 150, loss: 0.00432399520650506
step: 160, loss: 0.0010068339761346579
step: 170, loss: 0.0022098917979747057
step: 180, loss: 0.030919240787625313
step: 190, loss: 0.06423794478178024
step: 200, loss: 0.1650339812040329
step: 210, loss: 0.07941235601902008
/usr/local/lib/python3.7/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=0.5835010060362175, f1=0.5979797979797981,
best_f1=0.6041275797373358
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014706147834658623
step: 10, loss: 0.002549778902903199
step: 20, loss: 0.06043839827179909
step: 30, loss: 0.006866756360977888
step: 40, loss: 0.03465473651885986
step: 50, loss: 0.0009907607454806566
step: 60, loss: 0.00030192991835065186
step: 70, loss: 0.002913218457251787
step: 80, loss: 7.049436680972576e-05
step: 90, loss: 0.03885626047849655
step: 100, loss: 0.020349513739347458
step: 110, loss: 0.0011385283432900906

```

```

step: 120, loss: 0.001970811514183879
step: 130, loss: 0.0006409822963178158
step: 140, loss: 0.00047943854588083923
step: 150, loss: 0.0165509395301342
step: 160, loss: 0.007570717949420214
step: 170, loss: 0.0006123164785094559
step: 180, loss: 0.0005285930819809437
step: 190, loss: 0.031558990478515625
step: 200, loss: 0.07816347479820251
step: 210, loss: 0.003203953616321087
/usr/local/lib/python3.7/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=0.5910064239828693, f1=0.6020618556701032,
best_f1=0.6041275797373358
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0076276077888906
step: 10, loss: 0.015651654452085495
step: 20, loss: 0.0010922333458438516
step: 30, loss: 0.016625849530100822
step: 40, loss: 0.0018930042861029506
step: 50, loss: 0.0013819652376696467
step: 60, loss: 0.004301748238503933
step: 70, loss: 0.0028487059753388166
step: 80, loss: 0.003067572368308902
step: 90, loss: 0.0006492150714620948
step: 100, loss: 0.07819253206253052
step: 110, loss: 0.003393322927877307
step: 120, loss: 0.000350824004271999
step: 130, loss: 0.03215169906616211
step: 140, loss: 0.006497553084045649
step: 150, loss: 0.04341645538806915
step: 160, loss: 0.00018889624334406108
step: 170, loss: 0.07750257104635239
step: 180, loss: 0.0074019525200128555
step: 190, loss: 0.00212000566534698
step: 200, loss: 0.0003134555881842971
step: 210, loss: 0.00036431424086913466
/usr/local/lib/python3.7/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=0.5970149253731343, f1=0.5897435897435898,
best_f1=0.6041275797373358
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.012424662709236145
step: 10, loss: 0.0003982887137681246
step: 20, loss: 0.003909802529960871
step: 30, loss: 0.0011297132587060332
step: 40, loss: 0.018623139709234238
step: 50, loss: 0.01068457867950201
step: 60, loss: 0.000244682451011613
step: 70, loss: 0.0005945656448602676
step: 80, loss: 0.0007574179326184094
step: 90, loss: 0.00039228054811246693
step: 100, loss: 0.0005039922543801367
step: 110, loss: 0.0002775026368908584
step: 120, loss: 0.00030355234048329294
step: 130, loss: 0.001260837772861123
step: 140, loss: 0.0007856847951188684
step: 150, loss: 0.0002814109029714018
step: 160, loss: 0.0004982766695320606
step: 170, loss: 0.0051267254166305065
step: 180, loss: 0.00027016401872970164
step: 190, loss: 0.0018240216886624694
step: 200, loss: 0.10467977821826935
step: 210, loss: 0.0017373563023284078
/usr/local/lib/python3.7/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=0.5942857142857142, f1=0.6148007590132828,
best_f1=0.6041275797373358
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002728286199271679
step: 10, loss: 0.0005598802817985415
step: 20, loss: 0.005171294789761305
step: 30, loss: 0.004076115321367979
step: 40, loss: 0.005376377142965794
step: 50, loss: 0.0001960892550414428
step: 60, loss: 0.001303114928305149
step: 70, loss: 0.07886406034231186
step: 80, loss: 0.004793516825884581
step: 90, loss: 0.18079011142253876
step: 100, loss: 0.009274107404053211
step: 110, loss: 0.00017826490511652082
step: 120, loss: 0.00440913625061512
step: 130, loss: 0.00017938666860572994
step: 140, loss: 0.003962207119911909
step: 150, loss: 0.00018302162061445415
step: 160, loss: 0.0014965094160288572
step: 170, loss: 0.0004473027656786144

```

```

step: 180, loss: 0.008511475287377834
step: 190, loss: 0.000158999755512923
step: 200, loss: 0.0008815144537948072
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 210, loss: 0.0006794065702706575
/usr/local/lib/python3.7/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=0.5917926565874729, f1=0.5872340425531916,
best_f1=0.6041275797373358
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001519752200692892
step: 10, loss: 0.005322225857526064
step: 20, loss: 0.0001369473320664838
step: 30, loss: 0.0008249016827903688
step: 40, loss: 0.0018404403235763311
step: 50, loss: 0.0011798193445429206
step: 60, loss: 0.0029361362103372812
step: 70, loss: 0.01326133031398058
step: 80, loss: 0.020800532773137093
step: 90, loss: 0.00047081842785701156
step: 100, loss: 0.00027284593670628965
step: 110, loss: 0.00041905237594619393
step: 120, loss: 0.0030514420941472054
step: 130, loss: 0.0001465534296585247
step: 140, loss: 0.00038434515590779483
step: 150, loss: 0.009008634835481644
step: 160, loss: 0.00023596185201313347
step: 170, loss: 0.012551731429994106
step: 180, loss: 0.0019212007755413651
step: 190, loss: 0.0005003137630410492
step: 200, loss: 0.00020208551723044366
step: 210, loss: 0.0014690861571580172
/usr/local/lib/python3.7/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=0.5938144329896908, f1=0.6072874493927125,
best_f1=0.6041275797373358
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002043522661551833
step: 10, loss: 0.0007677703979425132
step: 20, loss: 0.024148333817720413
step: 30, loss: 0.00038677273551002145
step: 40, loss: 0.00018578051822260022

```



```

step: 50, loss: 0.00018175829609390348
step: 60, loss: 0.0004548237193375826
step: 70, loss: 0.0002547489129938185
step: 80, loss: 0.006113405339419842
step: 90, loss: 0.0002309204137418419
step: 100, loss: 0.00017991427739616483
step: 110, loss: 0.00037616171175614
step: 120, loss: 0.0003230970469303429
step: 130, loss: 0.0035335589200258255
step: 140, loss: 0.001165652647614479
step: 150, loss: 0.007658451795578003
step: 160, loss: 0.0002043378190137446
step: 170, loss: 0.0010860305046662688
step: 180, loss: 0.0003366238670423627
step: 190, loss: 0.00022282950521912426
step: 200, loss: 0.0009056829730980098
step: 210, loss: 0.01171758770942688
/usr/local/lib/python3.7/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=0.5862785862785863, f1=0.5967078189300411,
best_f1=0.6041275797373358
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

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

```

```

    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2293it [00:06, 344.17it/s]
load_f1 = 0.6150943396226415
real_f1 = 0.6139359698681732
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, 257.49it/s]

```

2.1.5 Walmart-Amazon - Running the matcher

```

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

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

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

```

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

Defaults for this optimization level are:

```

enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8621196150779724
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.16974873840808868
step: 20, loss: 0.1533418446779251
step: 30, loss: 0.5048061609268188
step: 40, loss: 0.25937268137931824
step: 50, loss: 0.3154776692390442
step: 60, loss: 0.36123862862586975
step: 70, loss: 0.17672735452651978
step: 80, loss: 0.5337506532669067
step: 90, loss: 0.23721976578235626
step: 100, loss: 0.22279970347881317
step: 110, loss: 0.23976954817771912
step: 120, loss: 0.42202073335647583
step: 130, loss: 0.3489396572113037
step: 140, loss: 0.32918933033943176

```

```

step: 150, loss: 0.265256404876709
step: 160, loss: 0.21378260850906372
step: 170, loss: 0.3909749984741211
step: 180, loss: 0.30171898007392883
step: 190, loss: 0.1210300475358963
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.41717791411042937, f1=0.43664717348927873,
best_f1=0.43664717348927873
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.30535951256752014
step: 10, loss: 0.04286384582519531
step: 20, loss: 0.0712834894657135
step: 30, loss: 0.14520035684108734
step: 40, loss: 0.46143075823783875
step: 50, loss: 0.33115342259407043
step: 60, loss: 0.1986061930656433
step: 70, loss: 0.21987342834472656
step: 80, loss: 0.12063029408454895
step: 90, loss: 0.17115093767642975
step: 100, loss: 0.24672631919384003
step: 110, loss: 0.12306917458772659
step: 120, loss: 0.36265018582344055
step: 130, loss: 0.17971840500831604
step: 140, loss: 0.22206948697566986
step: 150, loss: 0.04481670260429382
step: 160, loss: 0.1262086033821106
step: 170, loss: 0.23186177015304565
step: 180, loss: 0.20839115977287292
step: 190, loss: 0.22892315685749054
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7304347826086957, f1=0.7309941520467836,
best_f1=0.7309941520467836
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11244353652000427
step: 10, loss: 0.22639580070972443
step: 20, loss: 0.02436673454940319
step: 30, loss: 0.06513679772615433
step: 40, loss: 0.06107309088110924
step: 50, loss: 0.18114686012268066
step: 60, loss: 0.08139944821596146
step: 70, loss: 0.17177876830101013
step: 80, loss: 0.5014309287071228
step: 90, loss: 0.08051414787769318
step: 100, loss: 0.12247008085250854
step: 110, loss: 0.1988794207572937
step: 120, loss: 0.07866431027650833
step: 130, loss: 0.05923403427004814
step: 140, loss: 0.11202868074178696
step: 150, loss: 0.1431017518043518
step: 160, loss: 0.2098536491394043
step: 170, loss: 0.02513178437948227
step: 180, loss: 0.01181444339454174
step: 190, loss: 0.17413966357707977
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 3: dev_f1=0.7624309392265193, f1=0.7362637362637363,
best_f1=0.7362637362637363
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05616281181573868
step: 10, loss: 0.060600899159908295
step: 20, loss: 0.03372354060411453
step: 30, loss: 0.029128586873412132
step: 40, loss: 0.005814249627292156
step: 50, loss: 0.044193703681230545
step: 60, loss: 0.04132263734936714
step: 70, loss: 0.005015582777559757
step: 80, loss: 0.01685480773448944
step: 90, loss: 0.01389294397085905
step: 100, loss: 0.014004689641296864
step: 110, loss: 0.09134429693222046
step: 120, loss: 0.07873734831809998
step: 130, loss: 0.2934668958187103
step: 140, loss: 0.15657024085521698

```

```

step: 150, loss: 0.011633164249360561
step: 160, loss: 0.044295500963926315
step: 170, loss: 0.00995994359254837
step: 180, loss: 0.1689693033695221
step: 190, loss: 0.07217665016651154
/usr/local/lib/python3.7/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=0.736, f1=0.7621621621621621, best_f1=0.7362637362637363
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.16273446381092072
step: 10, loss: 0.015647875145077705
step: 20, loss: 0.21746954321861267
step: 30, loss: 0.04048115387558937
step: 40, loss: 0.07251288741827011
step: 50, loss: 0.1196337342262268
step: 60, loss: 0.018750159069895744
step: 70, loss: 0.006787164136767387
step: 80, loss: 0.004951062612235546
step: 90, loss: 0.02606183849275112
step: 100, loss: 0.034981198608875275
step: 110, loss: 0.0033165328204631805
step: 120, loss: 0.0004817927838303149
step: 130, loss: 0.013716673478484154
step: 140, loss: 0.019799506291747093
step: 150, loss: 0.0013386118225753307
step: 160, loss: 0.1268826425075531
step: 170, loss: 0.1146879568696022
step: 180, loss: 0.09713073074817657
step: 190, loss: 0.1340973824262619
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.775623268698061, f1=0.7621776504297996,
best_f1=0.7621776504297996
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05780033394694328
step: 10, loss: 0.0025007715448737144
step: 20, loss: 0.014107291586697102
step: 30, loss: 0.003025516401976347

```

```

step: 40, loss: 0.008540447801351547
step: 50, loss: 0.0036392107140272856
step: 60, loss: 0.0019945381209254265
step: 70, loss: 0.00527880759909749
step: 80, loss: 0.056867506355047226
step: 90, loss: 0.008885827846825123
step: 100, loss: 0.030517131090164185
step: 110, loss: 0.029804794117808342
step: 120, loss: 0.018129754811525345
step: 130, loss: 0.007069513667374849
step: 140, loss: 0.031034041196107864
step: 150, loss: 0.015376762486994267
step: 160, loss: 0.002262871479615569
step: 170, loss: 0.0635095164179802
step: 180, loss: 0.017639197409152985
step: 190, loss: 0.18677611649036407
/usr/local/lib/python3.7/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=0.7314285714285714, f1=0.7392550143266476,
best_f1=0.7621776504297996
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0027613788843154907
step: 10, loss: 0.006323775742202997
step: 20, loss: 0.0005525681190192699
step: 30, loss: 0.09619814157485962
step: 40, loss: 0.033623188734054565
step: 50, loss: 0.0404815673828125
step: 60, loss: 0.0031036229338496923
step: 70, loss: 0.0017178022535517812
step: 80, loss: 0.09484470635652542
step: 90, loss: 0.0020010373555123806
step: 100, loss: 0.002159417374059558
step: 110, loss: 0.008738021366298199
step: 120, loss: 0.0035029102582484484
step: 130, loss: 0.0008494448848068714
step: 140, loss: 0.0014898075023666024
step: 150, loss: 0.0018659074557945132
step: 160, loss: 0.011839581653475761
step: 170, loss: 0.0015818375395610929
step: 180, loss: 0.11356049031019211
step: 190, loss: 0.002869357354938984
/usr/local/lib/python3.7/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=0.7720207253886011, f1=0.7598944591029023,

```

```

best_f1=0.7621776504297996
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0032477304339408875
step: 10, loss: 0.02957111783325672
step: 20, loss: 0.00357392942532897
step: 30, loss: 0.03340214490890503
step: 40, loss: 0.018284006044268608
step: 50, loss: 0.05577594414353371
step: 60, loss: 0.02721349149942398
step: 70, loss: 0.0028575691394507885
step: 80, loss: 0.05075543746352196
step: 90, loss: 0.0021517565473914146
step: 100, loss: 0.01373310573399067
step: 110, loss: 0.004355635493993759
step: 120, loss: 0.0004454077279660851
step: 130, loss: 0.003755454905331135
step: 140, loss: 0.001454491401091218
step: 150, loss: 0.07469848543405533
step: 160, loss: 0.0018475569086149335
step: 170, loss: 0.00031135696917772293
step: 180, loss: 0.0011476044310256839
step: 190, loss: 0.009936746209859848
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7788944723618091, f1=0.760705289672544,
best_f1=0.760705289672544
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00483392458409071
step: 10, loss: 0.0011391034349799156
step: 20, loss: 0.003311350243166089
step: 30, loss: 0.009326450526714325
step: 40, loss: 0.07788460701704025
step: 50, loss: 0.0028971589636057615
step: 60, loss: 0.0003206531109753996
step: 70, loss: 0.0010810558451339602
step: 80, loss: 0.0021542347967624664
step: 90, loss: 0.001934607862494886
step: 100, loss: 0.0796930342912674
step: 110, loss: 0.0014912444166839123

```



```

step: 120, loss: 0.01397653017193079
step: 130, loss: 0.001004807185381651
step: 140, loss: 0.021237961947917938
step: 150, loss: 0.004222584888339043
step: 160, loss: 0.0010913352016359568
step: 170, loss: 0.02812916226685047
step: 180, loss: 0.012443268671631813
step: 190, loss: 0.021224360913038254
/usr/local/lib/python3.7/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=0.7684478371501272, f1=0.765498652291105,
best_f1=0.760705289672544
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005111293867230415
step: 10, loss: 0.00039762730011716485
step: 20, loss: 0.0003475596895441413
step: 30, loss: 0.00047968869330361485
step: 40, loss: 0.00031655761995352805
step: 50, loss: 0.0005296310409903526
step: 60, loss: 0.0013695042580366135
step: 70, loss: 0.0004919971106573939
step: 80, loss: 0.031870804727077484
step: 90, loss: 0.0023519624955952168
step: 100, loss: 0.00024259639030788094
step: 110, loss: 0.04035237058997154
step: 120, loss: 0.06538553535938263
step: 130, loss: 0.012922575697302818
step: 140, loss: 0.0004056008474435657
step: 150, loss: 0.0002006771246669814
step: 160, loss: 0.0010828387457877398
step: 170, loss: 0.000589642848353833
step: 180, loss: 0.0028161911759525537
step: 190, loss: 0.0006191874272190034
/usr/local/lib/python3.7/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=0.7493917274939174, f1=0.7587939698492463,
best_f1=0.760705289672544
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001405454007908702
step: 10, loss: 0.0005376967019401491
step: 20, loss: 0.0009238286293111742
step: 30, loss: 0.00026199090643785894

```

```

step: 40, loss: 0.0003385800519026816
step: 50, loss: 0.0006226477562449872
step: 60, loss: 0.0003948999219574034
step: 70, loss: 0.0002697932068258524
step: 80, loss: 0.0002556622202973813
step: 90, loss: 0.000249267352046445
step: 100, loss: 0.0005027563311159611
step: 110, loss: 0.001246858504600823
step: 120, loss: 0.01148857083171606
step: 130, loss: 0.00022017733135726303
step: 140, loss: 0.01639173924922943
step: 150, loss: 0.0003219651407562196
step: 160, loss: 0.0006593445432372391
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 170, loss: 0.0002791764563880861
step: 180, loss: 0.007819504477083683
step: 190, loss: 0.0007534088217653334
/usr/local/lib/python3.7/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=0.7611548556430445, f1=0.772117962466488,
best_f1=0.760705289672544
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00024527835194021463
step: 10, loss: 0.0007599437958560884
step: 20, loss: 0.03821473941206932
step: 30, loss: 0.0006357826641760767
step: 40, loss: 0.0005990301724523306
step: 50, loss: 0.00063986552413553
step: 60, loss: 0.0002920196857303381
step: 70, loss: 0.000391295732697472
step: 80, loss: 0.0002738413750194013
step: 90, loss: 0.00030440796399489045
step: 100, loss: 0.00016651647456455976
step: 110, loss: 0.011613705195486546
step: 120, loss: 0.00017027837748173624
step: 130, loss: 0.0001219615587615408
step: 140, loss: 0.0010228410828858614
step: 150, loss: 0.0002664015628397465
step: 160, loss: 0.02028210274875164
step: 170, loss: 0.0003940198803320527
step: 180, loss: 0.0007787655340507627
step: 190, loss: 0.0005749102565459907
/usr/local/lib/python3.7/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=0.7642679900744417, f1=0.7650000000000001,
best_f1=0.760705289672544
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00016330262587871403
step: 10, loss: 0.0010934752644971013
step: 20, loss: 0.0020240957383066416
step: 30, loss: 0.0004194152425043285
step: 40, loss: 0.00011369665298843756
step: 50, loss: 0.000854235258884728
step: 60, loss: 0.00029691067175008357
step: 70, loss: 0.001413271646015346
step: 80, loss: 0.00019117684860248119
step: 90, loss: 0.00405645789578557
step: 100, loss: 0.0003586566890589893
step: 110, loss: 0.001883413060568273
step: 120, loss: 0.0023918955121189356
step: 130, loss: 0.00028728711185976863
step: 140, loss: 0.0006153081194497645
step: 150, loss: 0.04450477287173271
step: 160, loss: 0.0003015877155121416
step: 170, loss: 0.00010573613690212369
step: 180, loss: 0.0012202938087284565
step: 190, loss: 0.0010655350051820278
/usr/local/lib/python3.7/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=0.7726027397260273, f1=0.7658402203856748,
best_f1=0.760705289672544
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0011215511476621032
step: 10, loss: 0.00032527869916521013
step: 20, loss: 0.0004496742330957204
step: 30, loss: 0.0013821111060678959
step: 40, loss: 9.748080628924072e-05
step: 50, loss: 0.05010980740189552
step: 60, loss: 0.00018713226018007845
step: 70, loss: 0.00019284624431747943
step: 80, loss: 0.0006100615719333291
step: 90, loss: 0.00018829981854651123
step: 100, loss: 0.0006817675894126296
step: 110, loss: 0.0004909602575935423
step: 120, loss: 0.0001802128681447357
step: 130, loss: 0.00012287554272916168
step: 140, loss: 0.00020716276776511222

```

```

step: 150, loss: 0.00040795261156745255
step: 160, loss: 0.00010508721607038751
step: 170, loss: 0.0001075038017006591
step: 180, loss: 0.0012943627079948783
step: 190, loss: 8.211790554923937e-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=0.7754010695187165, f1=0.7717391304347826,
best_f1=0.760705289672544
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001586221915204078
step: 10, loss: 6.0160920838825405e-05
step: 20, loss: 0.0002703167265281081
step: 30, loss: 0.00022669877216685563
step: 40, loss: 0.00026092701591551304
step: 50, loss: 0.00011336110037518665
step: 60, loss: 0.00016153235628735274
step: 70, loss: 0.0002609320217743516
step: 80, loss: 0.00019326676556374878
step: 90, loss: 0.001828828128054738
step: 100, loss: 0.0001830655528465286
step: 110, loss: 0.0001956722844624892
step: 120, loss: 0.0004109945730306208
step: 130, loss: 0.0005372174200601876
step: 140, loss: 0.0034628475550562143
step: 150, loss: 0.0003823896695394069
step: 160, loss: 0.0009304184932261705
step: 170, loss: 0.0024052767548710108
step: 180, loss: 0.00027202104683965445
step: 190, loss: 0.0006828055484220386
/usr/local/lib/python3.7/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=0.7754010695187165, f1=0.772972972972973,
best_f1=0.760705289672544
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

```

```

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True

```

```

master_weights      : True
loss_scale          : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled             : True
opt_level           : 02
cast_model_type     : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights      : True
loss_scale          : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2049it [00:09, 224.90it/s]
load_f1 = 0.7774936061381075
real_f1 = 0.7684478371501272
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, 256.28it/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: 0.8504493236541748
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.22049438953399658
step: 20, loss: 0.15354813635349274
step: 30, loss: 0.25554534792900085
step: 40, loss: 0.30184969305992126
step: 50, loss: 0.37793177366256714
step: 60, loss: 0.4388048052787781
step: 70, loss: 0.30499792098999023
step: 80, loss: 0.2569674551486969
step: 90, loss: 0.39711543917655945
step: 100, loss: 0.23006571829319
step: 110, loss: 0.17812590301036835
step: 120, loss: 0.5587731599807739
step: 130, loss: 0.41465476155281067
step: 140, loss: 0.4721519947052002
step: 150, loss: 0.11034049093723297
step: 160, loss: 0.30602553486824036
step: 170, loss: 0.14409597218036652
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6030927835051547, f1=0.6, best_f1=0.6
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.43537187576293945
step: 10, loss: 0.14038144052028656
step: 20, loss: 0.3118978440761566
step: 30, loss: 0.24094781279563904
step: 40, loss: 0.13109427690505981
step: 50, loss: 0.1955493986606598
step: 60, loss: 0.07469028234481812
step: 70, loss: 0.256046861410141
step: 80, loss: 0.10408519953489304
step: 90, loss: 0.26602718234062195
step: 100, loss: 0.17733463644981384
step: 110, loss: 0.18046370148658752
step: 120, loss: 0.04656171053647995
step: 130, loss: 0.08609098941087723
step: 140, loss: 0.08998888731002808

```

```

step: 150, loss: 0.18139773607254028
step: 160, loss: 0.05804609879851341
step: 170, loss: 0.05803443491458893
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7334963325183375, f1=0.7386091127098322,
best_f1=0.7386091127098322
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09250397235155106
step: 10, loss: 0.04007352516055107
step: 20, loss: 0.1525283008813858
step: 30, loss: 0.292548805475235
step: 40, loss: 0.011139369569718838
step: 50, loss: 0.07506552338600159
step: 60, loss: 0.0678059458732605
step: 70, loss: 0.07002535462379456
step: 80, loss: 0.21298961341381073
step: 90, loss: 0.05929568409919739
step: 100, loss: 0.04283164069056511
step: 110, loss: 0.07986010611057281
step: 120, loss: 0.008258940652012825
step: 130, loss: 0.12396399676799774
step: 140, loss: 0.021043069660663605
step: 150, loss: 0.19577060639858246
step: 160, loss: 0.020101655274629593
step: 170, loss: 0.17249581217765808
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7764127764127765, f1=0.7846889952153109,
best_f1=0.7846889952153109
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04999309405684471
step: 10, loss: 0.04801441729068756
step: 20, loss: 0.03309762105345726

```



```

step: 30, loss: 0.02599787898361683
step: 40, loss: 0.07639715075492859
step: 50, loss: 0.019063476473093033
step: 60, loss: 0.29705509543418884
step: 70, loss: 0.012841513380408287
step: 80, loss: 0.04469248652458191
step: 90, loss: 0.07725285738706589
step: 100, loss: 0.03646662458777428
step: 110, loss: 0.02837388589978218
step: 120, loss: 0.1356004774570465
step: 130, loss: 0.051459766924381256
step: 140, loss: 0.002640101360157132
step: 150, loss: 0.0032761108595877886
step: 160, loss: 0.022924885153770447
step: 170, loss: 0.014298536814749241
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.78125, f1=0.7564102564102565, best_f1=0.7564102564102565
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.015199704095721245
step: 10, loss: 0.07131636887788773
step: 20, loss: 0.013761098496615887
step: 30, loss: 0.024472670629620552
step: 40, loss: 0.007293174043297768
step: 50, loss: 0.004230526741594076
step: 60, loss: 0.002150158630684018
step: 70, loss: 0.027514148503541946
step: 80, loss: 0.012707289308309555
step: 90, loss: 0.1117108017206192
step: 100, loss: 0.04616628959774971
step: 110, loss: 0.009839790873229504
step: 120, loss: 0.1125757023692131
step: 130, loss: 0.0018702232046052814
step: 140, loss: 0.030437422916293144
step: 150, loss: 0.011607019230723381
step: 160, loss: 0.009385284967720509
step: 170, loss: 0.007446334231644869
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:

```

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=0.7837150127226464, f1=0.8048780487804877,
best_f1=0.8048780487804877
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.011007100343704224
step: 10, loss: 0.07512152194976807
step: 20, loss: 0.029807573184370995
step: 30, loss: 0.02902899496257305
step: 40, loss: 0.04115269333124161
step: 50, loss: 0.01761653833091259
step: 60, loss: 0.0036574983969330788
step: 70, loss: 0.00576270604506135
step: 80, loss: 0.026627451181411743
step: 90, loss: 0.004976087249815464
step: 100, loss: 0.013506581075489521
step: 110, loss: 0.02433919347822666
step: 120, loss: 0.003448310075327754
step: 130, loss: 0.24435201287269592
step: 140, loss: 0.0020158463157713413
step: 150, loss: 0.13626043498516083
step: 160, loss: 0.033839479088783264
step: 170, loss: 0.001211738446727395
/usr/local/lib/python3.7/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=0.7785888077858881, f1=0.7971014492753623,
best_f1=0.8048780487804877
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
```

```
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005414362531155348
step: 10, loss: 0.002583181019872427
step: 20, loss: 0.04059811308979988
step: 30, loss: 0.00556879211217165
step: 40, loss: 0.00404949439689517
step: 50, loss: 0.000714900263119489
step: 60, loss: 0.09434464573860168
step: 70, loss: 0.001794995041564107
step: 80, loss: 0.0011684667551890016
step: 90, loss: 0.012396853417158127
step: 100, loss: 0.10121176391839981
step: 110, loss: 0.0004205043660476804
step: 120, loss: 0.04305250197649002
step: 130, loss: 0.09230982512235641
```

```

step: 140, loss: 0.05461842566728592
step: 150, loss: 0.07028260082006454
step: 160, loss: 0.013543897308409214
step: 170, loss: 0.06290991604328156
/usr/local/lib/python3.7/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=0.7669172932330827, f1=0.7729468599033817,
best_f1=0.8048780487804877
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007215974386781454
step: 10, loss: 0.004752350505441427
step: 20, loss: 0.0019079898484051228
step: 30, loss: 0.05447756499052048
step: 40, loss: 0.018859364092350006
step: 50, loss: 0.0031874487176537514
step: 60, loss: 0.002481288742274046
step: 70, loss: 0.019673209637403488
step: 80, loss: 0.02017962373793125
step: 90, loss: 0.11041628569364548
step: 100, loss: 0.0008797704940661788
step: 110, loss: 0.013206020928919315
step: 120, loss: 0.0005969238118268549
step: 130, loss: 0.0005991225480102003
step: 140, loss: 0.0016740091377869248
step: 150, loss: 0.07315535098314285
step: 160, loss: 0.015813399106264114
step: 170, loss: 0.007779409177601337
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7888631090487238, f1=0.7643020594965675,
best_f1=0.7643020594965675
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0023232484236359596
step: 10, loss: 0.003349840408191085
step: 20, loss: 0.0005410861922428012
step: 30, loss: 0.022801123559474945
step: 40, loss: 0.009151380509138107
step: 50, loss: 0.02150236815214157

```

```

step: 60, loss: 0.02427561953663826
step: 70, loss: 0.002976709511131048
step: 80, loss: 0.01697668991982937
step: 90, loss: 0.03058188036084175
step: 100, loss: 0.0268703643232584
step: 110, loss: 9.125506767304614e-05
step: 120, loss: 0.004877832252532244
step: 130, loss: 0.0009067439241334796
step: 140, loss: 0.0006298914086073637
step: 150, loss: 0.019808337092399597
step: 160, loss: 0.012740460224449635
step: 170, loss: 0.003920922987163067
/usr/local/lib/python3.7/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=0.7804878048780487, f1=0.7616822429906542,
best_f1=0.7643020594965675
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0019310060888528824
step: 10, loss: 0.02362450212240219
step: 20, loss: 0.0002831158635672182
step: 30, loss: 0.0002188884827774018
step: 40, loss: 0.018137313425540924
step: 50, loss: 0.009162242524325848
step: 60, loss: 0.0039983391761779785
step: 70, loss: 0.00033971897209994495
step: 80, loss: 0.0008930182666517794
step: 90, loss: 0.02105453424155712
step: 100, loss: 0.04896995797753334
step: 110, loss: 0.00048258077003993094
step: 120, loss: 0.009758111089468002
step: 130, loss: 0.0018614813452586532
step: 140, loss: 0.008026355877518654
step: 150, loss: 0.00015366444131359458
step: 160, loss: 6.61133963149041e-05
step: 170, loss: 0.0001573527988512069
/usr/local/lib/python3.7/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=0.7802469135802469, f1=0.7912621359223301,
best_f1=0.7643020594965675
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00016458430036436766
step: 10, loss: 0.029615232720971107

```

```

step: 20, loss: 0.04321015998721123
step: 30, loss: 0.006973646581172943
step: 40, loss: 0.025276657193899155
step: 50, loss: 0.0003239162324462086
step: 60, loss: 0.0006345135043375194
step: 70, loss: 0.005334446672350168
step: 80, loss: 0.00166892574634403
step: 90, loss: 0.0028245393186807632
step: 100, loss: 0.0002732871798798442
step: 110, loss: 0.003676842665299773
step: 120, loss: 0.02680441364645958
step: 130, loss: 0.015774883329868317
step: 140, loss: 0.0677560567855835
step: 150, loss: 0.06365393102169037
step: 160, loss: 0.0027375936042517424
step: 170, loss: 0.25080204010009766
/usr/local/lib/python3.7/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=0.7733990147783252, f1=0.7872860635696822,
best_f1=0.7643020594965675
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.026785774156451225
step: 10, loss: 0.09541735798120499
step: 20, loss: 0.024349713698029518
step: 30, loss: 0.02375946007668972
step: 40, loss: 0.0002560988359618932
step: 50, loss: 0.005344816017895937
step: 60, loss: 0.0006864898605272174
step: 70, loss: 0.01934754103422165
step: 80, loss: 0.07377313077449799
step: 90, loss: 0.00025018485030159354
step: 100, loss: 0.01725638657808304
step: 110, loss: 0.0006166878738440573
step: 120, loss: 0.00037160320789553225
step: 130, loss: 0.0004185900907032192
step: 140, loss: 0.0012416592799127102
step: 150, loss: 0.0006518271402455866
step: 160, loss: 0.026977913454174995
step: 170, loss: 0.008489029482007027
/usr/local/lib/python3.7/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=0.7696078431372548, f1=0.7751196172248804,
best_f1=0.7643020594965675
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003137236926704645
step: 10, loss: 0.046973105520009995
step: 20, loss: 0.00046995541197247803
step: 30, loss: 0.00019146810518577695
step: 40, loss: 9.240122744813561e-05
step: 50, loss: 0.000987296923995018
step: 60, loss: 0.00014576595276594162
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 70, loss: 0.026539843529462814
step: 80, loss: 0.0007104146061465144
step: 90, loss: 0.0001641480776015669
step: 100, loss: 0.00011047568841604516
step: 110, loss: 0.0004964408581145108
step: 120, loss: 0.0003741441760212183
step: 130, loss: 0.00403082137927413
step: 140, loss: 0.0014961713459342718
step: 150, loss: 0.00020957448577973992
step: 160, loss: 0.00048026509466581047
step: 170, loss: 0.003484841901808977
/usr/local/lib/python3.7/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=0.7788461538461539, f1=0.7643020594965675,
best_f1=0.7643020594965675
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001966107520274818
step: 10, loss: 8.543298463337123e-05
step: 20, loss: 0.0012739785015583038
step: 30, loss: 0.004421962890774012
step: 40, loss: 0.005256367847323418
step: 50, loss: 0.0004495371540542692
step: 60, loss: 0.00012952415272593498
step: 70, loss: 0.0011364895617589355
step: 80, loss: 0.0016655392246320844
step: 90, loss: 0.0005407255957834423
step: 100, loss: 0.00034627006971277297
step: 110, loss: 0.0013529681600630283
step: 120, loss: 0.00011847306450363249
step: 130, loss: 0.00039356769411824644
step: 140, loss: 0.0014320387272164226
step: 150, loss: 0.0012299335794523358
step: 160, loss: 0.022522645071148872
step: 170, loss: 0.0019142619566991925
/usr/local/lib/python3.7/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=0.7772277227722771, f1=0.7846889952153109,
best_f1=0.7643020594965675
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0014685848727822304
step: 10, loss: 0.007489427924156189
step: 20, loss: 0.0004954221658408642
step: 30, loss: 0.02233896404504776
step: 40, loss: 9.724835399538279e-05
step: 50, loss: 0.002605739049613476
step: 60, loss: 8.026639989111573e-05
step: 70, loss: 0.0001853511348599568
step: 80, loss: 0.0009226236725226045
step: 90, loss: 8.51741642691195e-05
step: 100, loss: 0.00010570518497843295
step: 110, loss: 0.00030931239598430693
step: 120, loss: 0.002981101628392935
step: 130, loss: 0.00018144679779652506
step: 140, loss: 0.00028072012355551124
step: 150, loss: 0.04196002334356308
step: 160, loss: 7.21224641893059e-05
step: 170, loss: 0.006027492694556713
/usr/local/lib/python3.7/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=0.7791563275434243, f1=0.7846889952153109,
best_f1=0.7643020594965675
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

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

```

```

keep_batchnorm_fp32      : True
master_weights           : True
loss_scale               : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
1916it [00:04, 433.66it/s]
load_f1 = 0.7855421686746988
real_f1 = 0.7817745803357313
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, 339.44it/s]

```

2.3 BASELINE DIRTY

2.3.1 DBLP-ACM - Running the matcher

```

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

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

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

```



```

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

```

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

Defaults for this optimization level are:

```

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

```

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

After processing overrides, optimization options are:

```

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

```

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

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

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

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

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

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

```

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

```

```
step: 0, loss: 0.8161117434501648
```

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

```
step: 10, loss: 0.4785904586315155
```

```
step: 20, loss: 0.5843357443809509
```

```
step: 30, loss: 0.4582974910736084
```

```

step: 40, loss: 0.24111837148666382
step: 50, loss: 0.11118927597999573
step: 60, loss: 0.23501308262348175
step: 70, loss: 0.23011383414268494
step: 80, loss: 0.1901600956916809
step: 90, loss: 0.12753401696681976
step: 100, loss: 0.09979137778282166
step: 110, loss: 0.0801294595003128
step: 120, loss: 0.0303022637963295
step: 130, loss: 0.009498031809926033
step: 140, loss: 0.006752386689186096
step: 150, loss: 0.12000977247953415
step: 160, loss: 0.14927029609680176
step: 170, loss: 0.030271055176854134
step: 180, loss: 0.014972403645515442
step: 190, loss: 0.016296181827783585
step: 200, loss: 0.005573426838964224
step: 210, loss: 0.009824930690228939
step: 220, loss: 0.011593319475650787
step: 230, loss: 0.030721060931682587
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9497816593886462, f1=0.9524861878453039,
best_f1=0.9524861878453039
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08676332235336304
step: 10, loss: 0.02964177168905735
step: 20, loss: 0.004072367679327726
step: 30, loss: 0.00558505579829216
step: 40, loss: 0.030490977689623833
step: 50, loss: 0.060792408883571625
step: 60, loss: 0.006559300236403942
step: 70, loss: 0.014337769709527493
step: 80, loss: 0.017294341698288918
step: 90, loss: 0.010332894511520863
step: 100, loss: 0.12333761900663376
step: 110, loss: 0.06268981844186783
step: 120, loss: 0.027643820270895958
step: 130, loss: 0.015616883523762226
step: 140, loss: 0.17213556170463562
step: 150, loss: 0.010349179618060589

```

```

step: 160, loss: 0.0662086233496666
step: 170, loss: 0.1935110092163086
step: 180, loss: 0.06058798357844353
step: 190, loss: 0.05934485048055649
step: 200, loss: 0.026183659210801125
step: 210, loss: 0.10206770896911621
step: 220, loss: 0.0021504880860447884
step: 230, loss: 0.028368081897497177
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.967452300785634, f1=0.9686800894854586,
best_f1=0.9686800894854586
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.057254329323768616
step: 10, loss: 0.01766498014330864
step: 20, loss: 0.009331237524747849
step: 30, loss: 0.03586240112781525
step: 40, loss: 0.11093027889728546
step: 50, loss: 0.009171304292976856
step: 60, loss: 0.013884360902011395
step: 70, loss: 0.04026320204138756
step: 80, loss: 0.1684202402830124
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 90, loss: 0.09856563061475754
step: 100, loss: 0.014961316250264645
step: 110, loss: 0.01302423793822527
step: 120, loss: 0.002462438540533185
step: 130, loss: 0.0007462409557774663
step: 140, loss: 0.0010866179363802075
step: 150, loss: 0.005299133248627186
step: 160, loss: 0.0011364202946424484
step: 170, loss: 0.013589981943368912
step: 180, loss: 0.004595056641846895
step: 190, loss: 0.005925429053604603
step: 200, loss: 0.00714257312938571
step: 210, loss: 0.0018802640261128545
step: 220, loss: 0.012516900897026062
step: 230, loss: 0.06545420736074448
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9700996677740864, f1=0.9690265486725664,
best_f1=0.9690265486725664
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008376359939575195
step: 10, loss: 0.0013836082071065903
step: 20, loss: 0.0011983849108219147
step: 30, loss: 0.00896845106035471
step: 40, loss: 0.003607005812227726
step: 50, loss: 0.0027953439857810736
step: 60, loss: 0.002222184557467699
step: 70, loss: 0.006871899124234915
step: 80, loss: 0.15009154379367828
step: 90, loss: 0.008033600635826588
step: 100, loss: 0.0013746151234954596
step: 110, loss: 0.022381717339158058
step: 120, loss: 0.0028667214792221785
step: 130, loss: 0.044592976570129395
step: 140, loss: 0.0021378889214247465
step: 150, loss: 0.005039566662162542
step: 160, loss: 0.0024097964633256197
step: 170, loss: 0.0014197230339050293
step: 180, loss: 0.1935650259256363
step: 190, loss: 0.006678571458905935
step: 200, loss: 0.012706116773188114
step: 210, loss: 0.003002411685883999
step: 220, loss: 0.002558171981945634
step: 230, loss: 0.0010584512492641807
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9729119638826186, f1=0.9718785151856018,
best_f1=0.9718785151856018
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007078421767801046
step: 10, loss: 0.0043883007019758224
step: 20, loss: 0.0009681902593001723

```

```

step: 30, loss: 0.0005163777968846262
step: 40, loss: 0.0004622112901415676
step: 50, loss: 0.0008699627942405641
step: 60, loss: 0.0009959338931366801
step: 70, loss: 0.001427140785381198
step: 80, loss: 0.0012873366940766573
step: 90, loss: 0.0004419536853674799
step: 100, loss: 0.0020686709322035313
step: 110, loss: 0.000454584660474211
step: 120, loss: 0.06938467919826508
step: 130, loss: 0.04119883477687836
step: 140, loss: 0.0005460706888698041
step: 150, loss: 0.00024135470448527485
step: 160, loss: 0.11669471114873886
step: 170, loss: 0.013319388963282108
step: 180, loss: 0.005227569490671158
step: 190, loss: 0.001038739806972444
step: 200, loss: 0.036960966885089874
step: 210, loss: 0.006236034445464611
step: 220, loss: 0.03507228195667267
step: 230, loss: 0.0007484900415875018
/usr/local/lib/python3.7/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=0.9698996655518396, f1=0.9665924276169264,
best_f1=0.9718785151856018
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005939642433077097
step: 10, loss: 0.002572141820564866
step: 20, loss: 0.01849622093141079
step: 30, loss: 0.0026896181516349316
step: 40, loss: 0.0005438279476948082
step: 50, loss: 0.011121192015707493
step: 60, loss: 0.003032668260857463
step: 70, loss: 0.0007943689706735313
step: 80, loss: 0.0019923921208828688
step: 90, loss: 0.0025987294502556324
step: 100, loss: 0.002464042976498604
step: 110, loss: 0.02658151462674141
step: 120, loss: 0.0015416770474985242
step: 130, loss: 0.000508207711391151
step: 140, loss: 0.04072795435786247
step: 150, loss: 0.004999252501875162
step: 160, loss: 0.020596696063876152
step: 170, loss: 0.00022330516367219388
step: 180, loss: 0.0006340365507639945

```

```

step: 190, loss: 0.0012001466238871217
step: 200, loss: 0.0015717936912551522
step: 210, loss: 0.0004846287192776799
step: 220, loss: 0.0001505547552369535
step: 230, loss: 0.0013334190007299185
/usr/local/lib/python3.7/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=0.9663677130044843, f1=0.9731543624161074,
best_f1=0.9718785151856018
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.043624397367239
step: 10, loss: 0.00012903142487630248
step: 20, loss: 0.00033049657940864563
step: 30, loss: 0.00022961580543778837
step: 40, loss: 0.056382935494184494
step: 50, loss: 0.0006802050629630685
step: 60, loss: 0.04534679278731346
step: 70, loss: 0.00048426989815197885
step: 80, loss: 0.00010458407632540911
step: 90, loss: 0.0008604071335867047
step: 100, loss: 0.002664659870788455
step: 110, loss: 0.00016780479927547276
step: 120, loss: 0.00031493461574427783
step: 130, loss: 0.00027069999487139285
step: 140, loss: 9.253364987671375e-05
step: 150, loss: 0.00041395003790967166
step: 160, loss: 0.0001164758505183272
step: 170, loss: 0.00031249935273081064
step: 180, loss: 0.00018335727509111166
step: 190, loss: 0.0002990625216625631
step: 200, loss: 0.0025554420426487923
step: 210, loss: 0.000552992569282651
step: 220, loss: 0.0021276683546602726
step: 230, loss: 0.032680779695510864
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9752808988764046, f1=0.9720044792833147,
best_f1=0.9720044792833147
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.024569761008024216
step: 10, loss: 0.035200607031583786
step: 20, loss: 0.0003700953093357384
step: 30, loss: 0.00652864808216691
step: 40, loss: 0.004393712617456913
step: 50, loss: 0.0014158847043290734
step: 60, loss: 0.0007239841506816447
step: 70, loss: 0.0001417307648807764
step: 80, loss: 0.001954746665433049
step: 90, loss: 0.00010961145017063245
step: 100, loss: 0.00012562613119371235
step: 110, loss: 0.0003555155999492854
step: 120, loss: 0.00047873237053863704
step: 130, loss: 0.0010756540577858686
step: 140, loss: 0.004916804376989603
step: 150, loss: 0.0001656130189076066
step: 160, loss: 0.00036993634421378374
step: 170, loss: 0.02800111472606659
step: 180, loss: 0.00032567555899731815
step: 190, loss: 0.00033163660555146635
step: 200, loss: 0.005648649763315916
step: 210, loss: 0.00010919731721514836
step: 220, loss: 0.00040635248296894133
step: 230, loss: 0.015285419300198555
/usr/local/lib/python3.7/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=0.9695603156708005, f1=0.967525195968645,
best_f1=0.9720044792833147
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00031080664484761655
step: 10, loss: 0.0005413528415374458
step: 20, loss: 0.0004871104611083865
step: 30, loss: 0.00038799879257567227
step: 40, loss: 6.839801790192723e-05
step: 50, loss: 0.00020385460811667144
step: 60, loss: 0.0010193403577432036
step: 70, loss: 0.0005086258752271533
step: 80, loss: 0.054177768528461456
step: 90, loss: 0.02093185856938362
step: 100, loss: 0.0014891533646732569
step: 110, loss: 0.00013240022235549986
step: 120, loss: 0.1691100150346756
step: 130, loss: 0.00030578882433474064
step: 140, loss: 0.00024317955831065774

```

```

step: 150, loss: 6.750727334292606e-05
step: 160, loss: 0.0007114362670108676
step: 170, loss: 8.620421431260183e-05
step: 180, loss: 0.0007282022852450609
step: 190, loss: 0.0011829639552161098
step: 200, loss: 7.46545338188298e-05
step: 210, loss: 8.571326907258481e-05
step: 220, loss: 0.029192041605710983
step: 230, loss: 0.00043693569023162127
/usr/local/lib/python3.7/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=0.9717514124293786, f1=0.970917225950783,
best_f1=0.9720044792833147
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.959449430927634e-05
step: 10, loss: 4.902131695416756e-05
step: 20, loss: 8.145192259689793e-05
step: 30, loss: 0.0001171040566987358
step: 40, loss: 4.6315923100337386e-05
step: 50, loss: 0.0031577134504914284
step: 60, loss: 0.038374464958906174
step: 70, loss: 7.007297244854271e-05
step: 80, loss: 0.00011895886564161628
step: 90, loss: 0.006707132328301668
step: 100, loss: 7.353397813858464e-05
step: 110, loss: 6.431187648558989e-05
step: 120, loss: 0.02295977808535099
step: 130, loss: 0.00018250098219141364
step: 140, loss: 0.0009345689904876053
step: 150, loss: 0.00015913632523734123
step: 160, loss: 0.0001481335930293426
step: 170, loss: 0.00014312790881376714
step: 180, loss: 0.01225596759468317
step: 190, loss: 0.0001841739285737276
step: 200, loss: 7.958777860039845e-05
step: 210, loss: 0.0003179420600645244
step: 220, loss: 0.039941467344760895
step: 230, loss: 0.0001495133328717202
/usr/local/lib/python3.7/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=0.9743016759776536, f1=0.9698996655518396,
best_f1=0.9720044792833147
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```



```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00010567062417976558
step: 10, loss: 0.0002558029955253005
step: 20, loss: 5.634872286464088e-05
step: 30, loss: 8.145979518303648e-05
step: 40, loss: 0.005187165457755327
step: 50, loss: 0.00018547219224274158
step: 60, loss: 0.0003090042737312615
step: 70, loss: 0.00010468208347447217
step: 80, loss: 0.0001359646994387731
step: 90, loss: 0.0008379587670788169
step: 100, loss: 0.0004181162512395531
step: 110, loss: 6.15087483311072e-05
step: 120, loss: 0.0002081852435367182
step: 130, loss: 8.671118848724291e-05
step: 140, loss: 0.00016669703472871333
step: 150, loss: 4.3029169319197536e-05
step: 160, loss: 0.00010151587048312649
step: 170, loss: 0.02406669594347477
step: 180, loss: 0.00011579799320315942
step: 190, loss: 0.020844344049692154
step: 200, loss: 0.00020439928630366921
step: 210, loss: 9.113142732530832e-05
step: 220, loss: 0.0001937366760103032
step: 230, loss: 0.009957138448953629
/usr/local/lib/python3.7/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=0.9717514124293786, f1=0.9719416386083053,
best_f1=0.9720044792833147
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0011736948508769274
step: 10, loss: 0.0021203162614256144
step: 20, loss: 5.086167948320508e-05
step: 30, loss: 0.01355189923197031
step: 40, loss: 5.1851398893631995e-05
step: 50, loss: 6.506348290713504e-05
step: 60, loss: 0.010286131873726845
step: 70, loss: 0.00030264045926742256
step: 80, loss: 4.668882684200071e-05
step: 90, loss: 6.920856685610488e-05
step: 100, loss: 7.146714051486924e-05
step: 110, loss: 0.0001908496196847409
step: 120, loss: 0.00018720388470683247
step: 130, loss: 0.0005418396322056651
step: 140, loss: 5.231133400229737e-05

```

```

step: 150, loss: 0.00022922801144886762
step: 160, loss: 0.00024198443861678243
step: 170, loss: 0.00014412168820854276
step: 180, loss: 0.00010463297803653404
step: 190, loss: 0.02199334278702736
step: 200, loss: 0.003041475312784314
step: 210, loss: 0.00015617535973433405
step: 220, loss: 0.09263768047094345
step: 230, loss: 0.00011490984616102651
/usr/local/lib/python3.7/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=0.972972972972973,
best_f1=0.9720044792833147
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.595975769916549e-05
step: 10, loss: 0.000137213064590469
step: 20, loss: 0.00021084601758047938
step: 30, loss: 0.02538006566464901
step: 40, loss: 0.00020363263320177794
step: 50, loss: 0.00014252192340791225
step: 60, loss: 0.00010136916534975171
step: 70, loss: 0.0001570806634845212
step: 80, loss: 8.678658923599869e-05
step: 90, loss: 5.273878559819423e-05
step: 100, loss: 0.00022488120885100216
step: 110, loss: 0.00013945988030172884
step: 120, loss: 0.0004444062360562384
step: 130, loss: 0.0002835110353771597
step: 140, loss: 8.979371341411024e-05
step: 150, loss: 0.0001309648941969499
step: 160, loss: 9.935251728165895e-05
step: 170, loss: 7.106419798219576e-05
step: 180, loss: 0.0001270603679586202
step: 190, loss: 7.685721357120201e-05
step: 200, loss: 4.304762478568591e-05
step: 210, loss: 6.577213935088366e-05
step: 220, loss: 0.00020687771029770374
step: 230, loss: 5.1658047595992684e-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=0.9751131221719457, f1=0.9707865168539327,
best_f1=0.9720044792833147
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.115594391711056e-05
step: 10, loss: 3.5247088817413896e-05
step: 20, loss: 0.00018120359163731337
step: 30, loss: 9.72284033196047e-05
step: 40, loss: 5.648182923323475e-05
step: 50, loss: 7.095542969182134e-05
step: 60, loss: 0.0001104977200157009
step: 70, loss: 5.680323374690488e-05
step: 80, loss: 0.0002700173354241997
step: 90, loss: 0.00013634475180879235
step: 100, loss: 0.0016456878511235118
step: 110, loss: 0.00017488592129666358
step: 120, loss: 0.0001671940553933382
step: 130, loss: 9.321425022790208e-05
step: 140, loss: 0.00022894733410794288
step: 150, loss: 0.013141412287950516
step: 160, loss: 6.0071557527408004e-05
step: 170, loss: 6.453625246649608e-05
step: 180, loss: 0.00018584301869850606
step: 190, loss: 0.0025368172209709883
step: 200, loss: 4.938326310366392e-05
step: 210, loss: 0.00019121417426504195
step: 220, loss: 0.0035261823795735836
step: 230, loss: 3.3358730433974415e-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=0.9751693002257337, f1=0.9695603156708005,
best_f1=0.9720044792833147
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00012785660510417074
step: 10, loss: 7.153479964472353e-05
step: 20, loss: 0.0005529202171601355
step: 30, loss: 8.772116416366771e-05
step: 40, loss: 0.0002926778106484562
step: 50, loss: 7.737020496279001e-05
step: 60, loss: 4.276012987247668e-05
step: 70, loss: 6.926142668817192e-05
step: 80, loss: 0.00012855352542828768
step: 90, loss: 4.771871681441553e-05
step: 100, loss: 7.896297756815329e-05
step: 110, loss: 7.338010618695989e-05
step: 120, loss: 6.300039240159094e-05
step: 130, loss: 0.00013823708286508918
step: 140, loss: 0.002717576688155532

```

```

step: 150, loss: 0.00010056872270070016
step: 160, loss: 8.649026131024584e-05
step: 170, loss: 3.6908488254994154e-05
step: 180, loss: 6.037126149749383e-05
step: 190, loss: 0.006986710242927074
step: 200, loss: 4.076750701642595e-05
step: 210, loss: 0.009208577685058117
step: 220, loss: 0.0006235048058442771
step: 230, loss: 8.449858432868496e-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=0.9751693002257337, f1=0.9695603156708005,
best_f1=0.9720044792833147
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
2473it [00:08, 284.48it/s]
load_f1 = 0.9727891156462585
real_f1 = 0.971687429218573
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-

```

```
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:12, 354.51it/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        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
```

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

Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

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

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

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

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

```
step: 0, loss: 0.800672709941864
```

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

```
step: 10, loss: 0.4512139558792114
step: 20, loss: 0.508424699306488
step: 30, loss: 0.44271671772003174
step: 40, loss: 0.4191334843635559
step: 50, loss: 0.2518114745616913
step: 60, loss: 0.21384570002555847
step: 70, loss: 0.21465539932250977
step: 80, loss: 0.23447108268737793
step: 90, loss: 0.12921404838562012
step: 100, loss: 0.20822113752365112
step: 110, loss: 0.14315906167030334
step: 120, loss: 0.06323092430830002
step: 130, loss: 0.052347417920827866
step: 140, loss: 0.1176646426320076
step: 150, loss: 0.035491008311510086
step: 160, loss: 0.19157786667346954
step: 170, loss: 0.11353309452533722
step: 180, loss: 0.17141801118850708
step: 190, loss: 0.0625554695725441
step: 200, loss: 0.09437362104654312
step: 210, loss: 0.1452874094247818
step: 220, loss: 0.10229410231113434
```

```

step: 230, loss: 0.13204962015151978
step: 240, loss: 0.05457654222846031
step: 250, loss: 0.06857744604349136
step: 260, loss: 0.029505852609872818
step: 270, loss: 0.044433172792196274
step: 280, loss: 0.1545148640871048
step: 290, loss: 0.0727548897266388
step: 300, loss: 0.15682409703731537
step: 310, loss: 0.09932617098093033
step: 320, loss: 0.05992487072944641
step: 330, loss: 0.10535262525081635
step: 340, loss: 0.10145052522420883
step: 350, loss: 0.10002433508634567
step: 360, loss: 0.04678523540496826
step: 370, loss: 0.12485726922750473
step: 380, loss: 0.19753368198871613
step: 390, loss: 0.025605065748095512
step: 400, loss: 0.027028564363718033
step: 410, loss: 0.010542359203100204
step: 420, loss: 0.016531651839613914
step: 430, loss: 0.12398193776607513
step: 440, loss: 0.05838584527373314
step: 450, loss: 0.04226728528738022
step: 460, loss: 0.17328304052352905
step: 470, loss: 0.2791730463504791
step: 480, loss: 0.2918800115585327
step: 490, loss: 0.037257250398397446
step: 500, loss: 0.013002271763980389
step: 510, loss: 0.028071792796254158
step: 520, loss: 0.1432376652956009
step: 530, loss: 0.09118248522281647
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9207089552238805, f1=0.9173283512377393,
best_f1=0.9173283512377393
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11689640581607819
step: 10, loss: 0.1845160573720932
step: 20, loss: 0.16014644503593445
step: 30, loss: 0.06681234389543533
step: 40, loss: 0.008240253664553165

```

step: 50, loss: 0.018379155546426773
step: 60, loss: 0.11712218821048737
step: 70, loss: 0.2702394425868988
step: 80, loss: 0.01866730861365795
step: 90, loss: 0.001836372073739767
step: 100, loss: 0.2477894127368927
step: 110, loss: 0.03100394830107689
step: 120, loss: 0.07231531292200089
step: 130, loss: 0.029349083080887794
step: 140, loss: 0.034795600920915604
step: 150, loss: 0.03000817261636257
step: 160, loss: 0.10278698056936264
step: 170, loss: 0.13783538341522217
step: 180, loss: 0.010132263414561749
step: 190, loss: 0.0293789803981781
step: 200, loss: 0.02653200924396515
step: 210, loss: 0.02313670888543129
step: 220, loss: 0.15758520364761353
step: 230, loss: 0.07260321080684662
step: 240, loss: 0.1594492644071579
step: 250, loss: 0.08359086513519287
step: 260, loss: 0.011974141001701355
step: 270, loss: 0.1331184357404709
step: 280, loss: 0.2883850634098053
step: 290, loss: 0.12134214490652084
step: 300, loss: 0.034883320331573486
step: 310, loss: 0.08803076297044754
step: 320, loss: 0.20906822383403778
step: 330, loss: 0.01874607801437378
step: 340, loss: 0.003899351926520467
step: 350, loss: 0.06992290169000626
step: 360, loss: 0.013120713643729687
step: 370, loss: 0.014087148942053318
step: 380, loss: 0.06610427796840668
step: 390, loss: 0.020611489191651344
step: 400, loss: 0.012278109788894653
step: 410, loss: 0.0008462052792310715
step: 420, loss: 0.056700412184000015
step: 430, loss: 0.006690206006169319
step: 440, loss: 0.00545462965965271
step: 450, loss: 0.022149914875626564
step: 460, loss: 0.45484310388565063
step: 470, loss: 0.07180169969797134
step: 480, loss: 0.17895111441612244
step: 490, loss: 0.02291690558195114
step: 500, loss: 0.02155565842986107
step: 510, loss: 0.08024856448173523
step: 520, loss: 0.005206701811403036


```

step: 530, loss: 0.0734737291932106
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.930939226519337, f1=0.9269195189639223,
best_f1=0.9269195189639223
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003942043520510197
step: 10, loss: 0.026925021782517433
step: 20, loss: 0.16277310252189636
step: 30, loss: 0.25683650374412537
step: 40, loss: 0.0023341176565736532
step: 50, loss: 0.01761701889336109
step: 60, loss: 0.00450141541659832
step: 70, loss: 0.0636291578412056
step: 80, loss: 0.005947791039943695
step: 90, loss: 0.023126184940338135
step: 100, loss: 0.03412273898720741
step: 110, loss: 0.026008499786257744
step: 120, loss: 0.11708924174308777
step: 130, loss: 0.007867483422160149
step: 140, loss: 0.03758358210325241
step: 150, loss: 0.028802013024687767
step: 160, loss: 0.003165139351040125
step: 170, loss: 0.00392366386950016
step: 180, loss: 0.01021935697644949
step: 190, loss: 0.007754483725875616
step: 200, loss: 0.017952343448996544
step: 210, loss: 0.10781867057085037
step: 220, loss: 0.016136791557073593
step: 230, loss: 0.06711769849061966
step: 240, loss: 0.008862428367137909
step: 250, loss: 0.041579727083444595
step: 260, loss: 0.005057714879512787
step: 270, loss: 0.0033612994011491537
step: 280, loss: 0.019576558843255043
step: 290, loss: 0.026687413454055786
step: 300, loss: 0.03869916498661041
step: 310, loss: 0.12156261503696442
step: 320, loss: 0.1282639354467392
step: 330, loss: 0.011131527833640575
step: 340, loss: 0.019119160249829292

```

```

step: 350, loss: 0.007272642571479082
step: 360, loss: 0.009499283507466316
step: 370, loss: 0.018306706100702286
step: 380, loss: 0.08568976074457169
step: 390, loss: 0.02056676335632801
step: 400, loss: 0.07276707142591476
step: 410, loss: 0.017052240669727325
step: 420, loss: 0.06364009529352188
step: 430, loss: 0.04227149486541748
step: 440, loss: 0.027236711233854294
step: 450, loss: 0.0016589273000136018
step: 460, loss: 0.057778820395469666
step: 470, loss: 0.011523455381393433
step: 480, loss: 0.004462845157831907
step: 490, loss: 0.012633983045816422
step: 500, loss: 0.0743517130613327
step: 510, loss: 0.0036605193745344877
step: 520, loss: 0.014186165295541286
step: 530, loss: 0.08594334125518799
/usr/local/lib/python3.7/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=0.9245901639344263, f1=0.9242282507015902,
best_f1=0.9269195189639223
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0030850195325911045
step: 10, loss: 0.012853601947426796
step: 20, loss: 0.0047199903056025505
step: 30, loss: 0.002627057721838355
step: 40, loss: 0.06313776224851608
step: 50, loss: 0.02561407908797264
step: 60, loss: 0.0005247327499091625
step: 70, loss: 0.0024366809520870447
step: 80, loss: 0.043520718812942505
step: 90, loss: 0.015193368308246136
step: 100, loss: 0.008250473067164421
step: 110, loss: 0.0009246248519048095
step: 120, loss: 0.001384117640554905
step: 130, loss: 0.04812350496649742
step: 140, loss: 0.00848146341741085
step: 150, loss: 0.00980335008352995
step: 160, loss: 0.04304821789264679
step: 170, loss: 0.010555893182754517
step: 180, loss: 0.009902826510369778
step: 190, loss: 0.030112283304333687
step: 200, loss: 0.011161710135638714

```

```

step: 210, loss: 0.0017903604311868548
step: 220, loss: 0.0024990281090140343
step: 230, loss: 0.1661904901266098
step: 240, loss: 0.002358796773478389
step: 250, loss: 0.002164105884730816
step: 260, loss: 0.09127914160490036
step: 270, loss: 0.004365242552012205
step: 280, loss: 0.024259958416223526
step: 290, loss: 0.026481999084353447
step: 300, loss: 0.0004316759295761585
step: 310, loss: 0.009657077491283417
step: 320, loss: 0.003753487253561616
step: 330, loss: 0.029526440426707268
step: 340, loss: 0.09010016918182373
step: 350, loss: 0.0008046450675465167
step: 360, loss: 0.11405893415212631
step: 370, loss: 0.005848679691553116
step: 380, loss: 0.0028299225959926844
step: 390, loss: 0.008158006705343723
step: 400, loss: 0.012954216450452805
step: 410, loss: 0.005948287434875965
step: 420, loss: 0.002401350298896432
step: 430, loss: 0.006872020196169615
step: 440, loss: 0.17638671398162842
step: 450, loss: 0.01164548471570015
step: 460, loss: 0.0009586724336259067
step: 470, loss: 0.005325960926711559
step: 480, loss: 0.0009527301299385726
step: 490, loss: 0.0024499341379851103
step: 500, loss: 0.0025268997997045517
step: 510, loss: 0.0018264537211507559
step: 520, loss: 0.022320110350847244
step: 530, loss: 0.0018197862664237618
/usr/local/lib/python3.7/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=0.9282739472466451, f1=0.9203703703703704,
best_f1=0.9269195189639223
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0044912006705999374
step: 10, loss: 0.0025441306643188
step: 20, loss: 0.025508388876914978
step: 30, loss: 0.003151171375066042
step: 40, loss: 0.0026067199651151896
step: 50, loss: 0.033907532691955566
step: 60, loss: 0.02091604843735695

```

step: 70, loss: 0.0006850216304883361
step: 80, loss: 0.001656938693486154
step: 90, loss: 0.0009743777336552739
step: 100, loss: 0.0007650200859643519
step: 110, loss: 0.016234297305345535
step: 120, loss: 0.0017830007709562778
step: 130, loss: 0.0008366184774786234
step: 140, loss: 0.0005227820365689695
step: 150, loss: 0.0007773750694468617
step: 160, loss: 0.08407391607761383
step: 170, loss: 0.0025443295016884804
step: 180, loss: 0.004060927778482437
step: 190, loss: 0.0002691027766559273
step: 200, loss: 0.020246334373950958
step: 210, loss: 0.00047429310507141054
step: 220, loss: 0.000373541668523103
step: 230, loss: 0.0003795028605964035
step: 240, loss: 0.038185492157936096
step: 250, loss: 0.0003160550259053707
step: 260, loss: 0.0016029865946620703
step: 270, loss: 0.03676650673151016
step: 280, loss: 0.029917335137724876
step: 290, loss: 0.016144540160894394
step: 300, loss: 0.007709196303039789
step: 310, loss: 0.011168557219207287
step: 320, loss: 0.00477898633107543
step: 330, loss: 0.002634343458339572
step: 340, loss: 0.0010401050094515085
step: 350, loss: 0.0024419943802058697
step: 360, loss: 0.13725076615810394
step: 370, loss: 0.024882925674319267
step: 380, loss: 0.05243774875998497
step: 390, loss: 0.003537687472999096
step: 400, loss: 0.028019215911626816
step: 410, loss: 0.001062967348843813
step: 420, loss: 0.006550235208123922
step: 430, loss: 0.00567214610055089
step: 440, loss: 0.0007680029375478625
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 450, loss: 0.10716409236192703
step: 460, loss: 0.00512287812307477
step: 470, loss: 0.004244452342391014
step: 480, loss: 0.004071055445820093
step: 490, loss: 0.014210807159543037
step: 500, loss: 0.018303997814655304
step: 510, loss: 0.0019704506266862154
step: 520, loss: 0.0035347193479537964
step: 530, loss: 0.0014617698034271598

```
/usr/local/lib/python3.7/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=0.9304467987102718, f1=0.9285384970032272,
best_f1=0.9269195189639223
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0021110232919454575
step: 10, loss: 0.01742398925125599
step: 20, loss: 0.00453413650393486
step: 30, loss: 0.016885429620742798
step: 40, loss: 0.0038990573957562447
step: 50, loss: 0.0006661559455096722
step: 60, loss: 0.0004627892922144383
step: 70, loss: 0.20506331324577332
step: 80, loss: 0.0026979828253388405
step: 90, loss: 0.0007302975282073021
step: 100, loss: 0.001277638366445899
step: 110, loss: 0.0030648019164800644
step: 120, loss: 0.0014420161023736
step: 130, loss: 0.00038716712151654065
step: 140, loss: 0.00032688520150259137
step: 150, loss: 0.0011161514557898045
step: 160, loss: 0.00032241447479464114
step: 170, loss: 0.00029861045186407864
step: 180, loss: 0.022324195131659508
step: 190, loss: 0.002631671726703644
step: 200, loss: 0.00018394054495729506
step: 210, loss: 0.00013093669258523732
step: 220, loss: 0.0002138991403626278
step: 230, loss: 0.00018685508985072374
step: 240, loss: 0.00023518723901361227
step: 250, loss: 0.000370311870938167
step: 260, loss: 0.0006521135801449418
step: 270, loss: 0.0005609398940578103
step: 280, loss: 0.0011159913847222924
step: 290, loss: 0.10696981847286224
step: 300, loss: 0.013180826790630817
step: 310, loss: 0.012065442278981209
step: 320, loss: 0.04414477199316025
step: 330, loss: 0.002369540510699153
step: 340, loss: 0.007441291585564613
step: 350, loss: 0.023105613887310028
step: 360, loss: 0.0003003271121997386
step: 370, loss: 0.0012791093904525042
step: 380, loss: 0.004085305146872997
step: 390, loss: 0.0429135225713253
```

```

step: 400, loss: 0.001117478939704597
step: 410, loss: 0.000706857827026397
step: 420, loss: 0.0012176232412457466
step: 430, loss: 0.0007723108865320683
step: 440, loss: 0.01684650219976902
step: 450, loss: 0.001841370714828372
step: 460, loss: 0.0006547250086441636
step: 470, loss: 0.12099011987447739
step: 480, loss: 0.011005102656781673
step: 490, loss: 0.03614039719104767
step: 500, loss: 0.03530314937233925
step: 510, loss: 0.006133080460131168
step: 520, loss: 0.0218727458268404
step: 530, loss: 0.036493122577667236
/usr/local/lib/python3.7/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=0.903225806451613, f1=0.9014218009478674,
best_f1=0.9269195189639223
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01386907696723938
step: 10, loss: 0.006155775859951973
step: 20, loss: 0.11958260834217072
step: 30, loss: 0.0022564525716006756
step: 40, loss: 0.0005383477546274662
step: 50, loss: 0.0021106444764882326
step: 60, loss: 0.0022213924676179886
step: 70, loss: 0.0025033126585185528
step: 80, loss: 0.0009972050320357084
step: 90, loss: 0.002021425636485219
step: 100, loss: 0.0005437753861770034
step: 110, loss: 0.0054688844829797745
step: 120, loss: 0.00038408426917158067
step: 130, loss: 0.040738947689533234
step: 140, loss: 0.008363254368305206
step: 150, loss: 0.00020330780535005033
step: 160, loss: 0.0028065573424100876
step: 170, loss: 0.00023897213395684958
step: 180, loss: 0.0002620223385747522
step: 190, loss: 0.0004380344762466848
step: 200, loss: 0.0013544365065172315
step: 210, loss: 0.0008024830603972077
step: 220, loss: 0.0003222704981453717
step: 230, loss: 0.0004412556590978056
step: 240, loss: 0.0011263629421591759
step: 250, loss: 0.0011049159802496433

```

```

step: 260, loss: 0.148037850856781
step: 270, loss: 0.001083842129446566
step: 280, loss: 0.001159638981334865
step: 290, loss: 0.0047680712305009365
step: 300, loss: 0.0004430901608429849
step: 310, loss: 0.00034902876359410584
step: 320, loss: 0.002850852906703949
step: 330, loss: 0.00027497074916027486
step: 340, loss: 0.0026886777486652136
step: 350, loss: 0.01784830167889595
step: 360, loss: 0.0038047924172133207
step: 370, loss: 0.0008695591241121292
step: 380, loss: 0.0019085526000708342
step: 390, loss: 9.403418516740203e-05
step: 400, loss: 0.0004507877747528255
step: 410, loss: 0.0027136681601405144
step: 420, loss: 0.00015088044165167958
step: 430, loss: 7.973547326400876e-05
step: 440, loss: 0.00015332661860156804
step: 450, loss: 0.0009943107143044472
step: 460, loss: 0.001961770234629512
step: 470, loss: 0.0022138147614896297
step: 480, loss: 0.00023389853595290333
step: 490, loss: 0.0008669166127219796
step: 500, loss: 0.019986940547823906
step: 510, loss: 0.0018658069893717766
step: 520, loss: 0.00017868191935122013
step: 530, loss: 8.25066672405228e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 7: dev_f1=0.9211267605633803, f1=0.9200561009817672,
best_f1=0.9269195189639223
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006106981309130788
step: 10, loss: 0.001048276899382472
step: 20, loss: 0.05723497271537781
step: 30, loss: 0.0010343336034566164
step: 40, loss: 0.022875746712088585
step: 50, loss: 0.000869392475578934
step: 60, loss: 0.000527001335285604
step: 70, loss: 0.001066308468580246
step: 80, loss: 0.0005264205974526703
step: 90, loss: 0.0004316499689593911
step: 100, loss: 0.0016783189494162798
step: 110, loss: 0.0012074803235009313

```

```

step: 120, loss: 0.0004075938486494124
step: 130, loss: 0.003918763250112534
step: 140, loss: 6.473269604612142e-05
step: 150, loss: 0.0027986241038888693
step: 160, loss: 0.0004713160451501608
step: 170, loss: 0.00034774834057316184
step: 180, loss: 0.0010259286500513554
step: 190, loss: 0.0001380607282044366
step: 200, loss: 0.0012519042938947678
step: 210, loss: 0.00020676173153333366
step: 220, loss: 0.0014322915812954307
step: 230, loss: 0.0001548731088405475
step: 240, loss: 7.44564677006565e-05
step: 250, loss: 0.0006234686588868499
step: 260, loss: 0.04595019295811653
step: 270, loss: 5.6729360949248075e-05
step: 280, loss: 0.000195655768038705
step: 290, loss: 0.0008917040540836751
step: 300, loss: 0.00018047534103970975
step: 310, loss: 0.004946815315634012
step: 320, loss: 0.000923526065889746
step: 330, loss: 0.02781928889453411
step: 340, loss: 0.032961148768663406
step: 350, loss: 0.003865591250360012
step: 360, loss: 0.00011919924145331606
step: 370, loss: 0.0580277219414711
step: 380, loss: 0.00013391811808105558
step: 390, loss: 0.010780519805848598
step: 400, loss: 0.0034937956370413303
step: 410, loss: 0.000543405709322542
step: 420, loss: 0.00012438252451829612
step: 430, loss: 0.0005736147868447006
step: 440, loss: 0.07074091583490372
step: 450, loss: 0.0016137995989993215
step: 460, loss: 0.0011237752623856068
step: 470, loss: 0.008877037093043327
step: 480, loss: 0.0004766569472849369
step: 490, loss: 0.00022709420591127127
step: 500, loss: 0.00038018589839339256
step: 510, loss: 0.00014030825695954263
step: 520, loss: 0.00020230257359798998
step: 530, loss: 0.00028795181424356997
/usr/local/lib/python3.7/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=0.9292279411764707, f1=0.9247113163972287,
best_f1=0.9269195189639223
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```


An input tensor was not cuda.

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

```
step: 0, loss: 0.0012695675250142813
step: 10, loss: 0.00016508593398611993
step: 20, loss: 0.000117624229460489
step: 30, loss: 0.00022133156016934663
step: 40, loss: 0.00037077319575473666
step: 50, loss: 0.001739818137139082
step: 60, loss: 0.00017944886349141598
step: 70, loss: 0.11271639913320541
step: 80, loss: 0.00037904505734331906
step: 90, loss: 0.0002832900208886713
step: 100, loss: 0.00010528611164772883
step: 110, loss: 0.0005672459374181926
step: 120, loss: 0.00011591362999752164
step: 130, loss: 0.00018016727699432522
step: 140, loss: 0.0055949492380023
step: 150, loss: 6.352802301989868e-05
step: 160, loss: 7.957392517710105e-05
step: 170, loss: 9.855546522885561e-05
step: 180, loss: 5.3503776143770665e-05
step: 190, loss: 0.0011339862830936909
step: 200, loss: 0.00017583831504452974
step: 210, loss: 0.00018019853450823575
step: 220, loss: 0.0003998104075435549
step: 230, loss: 0.0040558152832090855
step: 240, loss: 0.0005723046488128603
step: 250, loss: 9.832771320361644e-05
step: 260, loss: 8.02980866865255e-05
step: 270, loss: 0.0001331967068836093
step: 280, loss: 0.000526233168784529
step: 290, loss: 6.980510079301894e-05
step: 300, loss: 0.0004657102399505675
step: 310, loss: 5.320714262779802e-05
step: 320, loss: 9.580834012012929e-05
step: 330, loss: 0.0010513376910239458
step: 340, loss: 8.508165046805516e-05
step: 350, loss: 7.727778574917465e-05
step: 360, loss: 0.004028765484690666
step: 370, loss: 0.00015105711645446718
step: 380, loss: 0.0003356098895892501
step: 390, loss: 0.0005794708267785609
step: 400, loss: 0.0006320676766335964
step: 410, loss: 0.00016141784726642072
step: 420, loss: 0.00040736395749263465
step: 430, loss: 6.26435357844457e-05
step: 440, loss: 0.1257089525461197
step: 450, loss: 0.00014053094491828233
```

```

step: 460, loss: 0.0001830875116866082
step: 470, loss: 0.00019505308591760695
step: 480, loss: 0.00020706283976323903
step: 490, loss: 0.00013788849173579365
step: 500, loss: 0.00014848774299025536
step: 510, loss: 0.0006663509993813932
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 520, loss: 0.00011607891792664304
step: 530, loss: 5.136823529028334e-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=0.9250234301780694, f1=0.9200945626477541,
best_f1=0.9269195189639223
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0028774673119187355
step: 10, loss: 6.759465031791478e-05
step: 20, loss: 0.00013129347644280642
step: 30, loss: 8.536796667613089e-05
step: 40, loss: 0.00018884355085901916
step: 50, loss: 9.250840957975015e-05
step: 60, loss: 0.0001153789708041586
step: 70, loss: 0.00017829281568992883
step: 80, loss: 0.003699499648064375
step: 90, loss: 0.00029313916456885636
step: 100, loss: 9.127323573920876e-05
step: 110, loss: 0.0014986463356763124
step: 120, loss: 0.00023825923562981188
step: 130, loss: 7.896820898167789e-05
step: 140, loss: 0.00041837102617137134
step: 150, loss: 0.00013451321865431964
step: 160, loss: 0.0003794910735450685
step: 170, loss: 0.0023079284001141787
step: 180, loss: 0.0013906934764236212
step: 190, loss: 0.0024608923122286797
step: 200, loss: 8.588148193666711e-05
step: 210, loss: 4.0652132156537846e-05
step: 220, loss: 5.399654764914885e-05
step: 230, loss: 0.00010483067308086902
step: 240, loss: 5.592970410361886e-05
step: 250, loss: 0.03410468250513077
step: 260, loss: 0.03065522387623787
step: 270, loss: 2.810230216709897e-05
step: 280, loss: 5.79644984100014e-05
step: 290, loss: 7.668229227419943e-05
step: 300, loss: 0.009427033364772797

```

```

step: 310, loss: 3.215163087588735e-05
step: 320, loss: 0.0004277856496628374
step: 330, loss: 0.00047089066356420517
step: 340, loss: 0.0003292947367299348
step: 350, loss: 4.864538641413674e-05
step: 360, loss: 0.0003549603861756623
step: 370, loss: 0.0008305193041451275
step: 380, loss: 6.640327774221078e-05
step: 390, loss: 0.1746005117893219
step: 400, loss: 0.00027131522074341774
step: 410, loss: 0.00019618267833720893
step: 420, loss: 0.010370655916631222
step: 430, loss: 8.762416837271303e-05
step: 440, loss: 8.898059604689479e-05
step: 450, loss: 0.009319315664470196
step: 460, loss: 0.004071695264428854
step: 470, loss: 7.710924546699971e-05
step: 480, loss: 0.0007081038784235716
step: 490, loss: 0.028439467772841454
step: 500, loss: 0.0009387219324707985
step: 510, loss: 0.00045094589586369693
step: 520, loss: 0.0017450524028390646
step: 530, loss: 0.0001208862813655287
/usr/local/lib/python3.7/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=0.9263456090651558, f1=0.9161475802587445,
best_f1=0.9269195189639223
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00037273604539223015
step: 10, loss: 0.0037420932203531265
step: 20, loss: 0.0007692137151025236
step: 30, loss: 4.416036244947463e-05
step: 40, loss: 0.0003971232217736542
step: 50, loss: 0.0009150226251222193
step: 60, loss: 0.0001279899588553235
step: 70, loss: 0.00014681137690786272
step: 80, loss: 0.012665924616158009
step: 90, loss: 0.00022237400116864592
step: 100, loss: 0.0002430529275443405
step: 110, loss: 0.00022809889924246818
step: 120, loss: 0.0007801146130077541
step: 130, loss: 0.0004131350724492222
step: 140, loss: 0.0019462836207821965
step: 150, loss: 0.00040166053804568946
step: 160, loss: 0.0008536295499652624

```

```

step: 170, loss: 0.0014331304701045156
step: 180, loss: 0.00010809423110913485
step: 190, loss: 0.0005735462182201445
step: 200, loss: 0.0010827461956068873
step: 210, loss: 7.562135579064488e-05
step: 220, loss: 0.0024332976900041103
step: 230, loss: 0.00013646938896272331
step: 240, loss: 0.0002992125228047371
step: 250, loss: 0.0001978942600544542
step: 260, loss: 0.00038478983333334327
step: 270, loss: 0.0012281823437660933
step: 280, loss: 0.0015125677455216646
step: 290, loss: 0.001432725926861167
step: 300, loss: 0.000538424588739872
step: 310, loss: 0.0010367392096668482
step: 320, loss: 0.0003016409755218774
step: 330, loss: 0.0013939954806119204
step: 340, loss: 8.573640661779791e-05
step: 350, loss: 0.09157680720090866
step: 360, loss: 7.806556823197752e-05
step: 370, loss: 5.581118239206262e-05
step: 380, loss: 0.0004461002827156335
step: 390, loss: 0.005410580895841122
step: 400, loss: 3.53661562257912e-05
step: 410, loss: 6.942241452634335e-05
step: 420, loss: 0.00016969302669167519
step: 430, loss: 0.011570672504603863
step: 440, loss: 8.49485004437156e-05
step: 450, loss: 5.6906737881945446e-05
step: 460, loss: 0.00018986855866387486
step: 470, loss: 0.001119549386203289
step: 480, loss: 0.0003627074183896184
step: 490, loss: 0.0004638328100554645
step: 500, loss: 4.6458477299893275e-05
step: 510, loss: 5.1548780902521685e-05
step: 520, loss: 6.439552089432254e-05
step: 530, loss: 3.6150879168417305e-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=0.9258215962441314, f1=0.9245372567631704,
best_f1=0.9269195189639223
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.4045740398578346e-05
step: 10, loss: 7.304656901396811e-05
step: 20, loss: 4.2367668356746435e-05

```

step: 30, loss: 5.635871275444515e-05
step: 40, loss: 0.00013928435510024428
step: 50, loss: 4.0398561395704746e-05
step: 60, loss: 4.777285357704386e-05
step: 70, loss: 0.00010253622895106673
step: 80, loss: 0.00015817198436707258
step: 90, loss: 8.497734233969823e-05
step: 100, loss: 0.00012646678078453988
step: 110, loss: 6.1018548876745626e-05
step: 120, loss: 8.867478027241305e-05
step: 130, loss: 4.4827040255768225e-05
step: 140, loss: 8.181517478078604e-05
step: 150, loss: 0.00024277472402900457
step: 160, loss: 4.125146733713336e-05
step: 170, loss: 5.725033406633884e-05
step: 180, loss: 0.00023554523068014532
step: 190, loss: 3.4362616133876145e-05
step: 200, loss: 5.560491990763694e-05
step: 210, loss: 8.350782445631921e-05
step: 220, loss: 3.374216612428427e-05
step: 230, loss: 5.105350646772422e-05
step: 240, loss: 7.507342525059357e-05
step: 250, loss: 0.0004634464858099818
step: 260, loss: 0.00011989823542535305
step: 270, loss: 0.00012479664292186499
step: 280, loss: 4.411331974552013e-05
step: 290, loss: 0.0003347531892359257
step: 300, loss: 0.00011437865032348782
step: 310, loss: 4.14559654018376e-05
step: 320, loss: 5.675295687979087e-05
step: 330, loss: 3.430056312936358e-05
step: 340, loss: 4.6786139137111604e-05
step: 350, loss: 0.007475986145436764
step: 360, loss: 0.0002863389381673187
step: 370, loss: 2.890743053285405e-05
step: 380, loss: 7.468323019566014e-05
step: 390, loss: 3.852380905300379e-05
step: 400, loss: 0.00021339226805139333
step: 410, loss: 0.015549424104392529
step: 420, loss: 0.00051043217536062
step: 430, loss: 3.766849476960488e-05
step: 440, loss: 6.070544259273447e-05
step: 450, loss: 0.006570253521203995
step: 460, loss: 5.1535509555833414e-05
step: 470, loss: 0.0003916625864803791
step: 480, loss: 0.00021467161423061043
step: 490, loss: 0.0005258776945993304
step: 500, loss: 0.0007178683299571276

```

step: 510, loss: 5.1664399506989866e-05
step: 520, loss: 0.00020780874183401465
step: 530, loss: 0.00017032383766490966
/usr/local/lib/python3.7/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=0.9280074314909429, f1=0.923581809657759,
best_f1=0.9269195189639223
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00039111380465328693
step: 10, loss: 3.106609437963925e-05
step: 20, loss: 9.061580203706399e-05
step: 30, loss: 0.00010686814493965358
step: 40, loss: 4.890644777333364e-05
step: 50, loss: 6.229138671187684e-05
step: 60, loss: 4.165606151218526e-05
step: 70, loss: 2.8448512239265256e-05
step: 80, loss: 4.078484926139936e-05
step: 90, loss: 6.969157402636483e-05
step: 100, loss: 2.049977774731815e-05
step: 110, loss: 4.8649180826032534e-05
step: 120, loss: 6.127567758085206e-05
step: 130, loss: 7.794945122441277e-05
step: 140, loss: 4.991846071789041e-05
step: 150, loss: 2.8319018383626826e-05
step: 160, loss: 2.740609852480702e-05
step: 170, loss: 3.226368789910339e-05
step: 180, loss: 2.9585324227809906e-05
step: 190, loss: 3.399193155928515e-05
step: 200, loss: 0.002156693022698164
step: 210, loss: 0.0014440567465499043
step: 220, loss: 2.456025504216086e-05
step: 230, loss: 4.0477381844539195e-05
step: 240, loss: 5.524131120182574e-05
step: 250, loss: 0.01990974135696888
step: 260, loss: 0.00012734413030557334
step: 270, loss: 6.362595013342798e-05
step: 280, loss: 7.704456220380962e-05
step: 290, loss: 5.1087510655634105e-05
step: 300, loss: 2.802448216243647e-05
step: 310, loss: 0.00010392443073214963
step: 320, loss: 3.37185847456567e-05
step: 330, loss: 8.303439244627953e-05
step: 340, loss: 2.9276165150804445e-05
step: 350, loss: 3.586639650166035e-05
step: 360, loss: 3.237010969314724e-05

```

```

step: 370, loss: 4.253830775269307e-05
step: 380, loss: 0.000265290291281417
step: 390, loss: 2.2891479602549225e-05
step: 400, loss: 2.3420176148647442e-05
step: 410, loss: 2.7316675186739303e-05
step: 420, loss: 2.5763511075638235e-05
step: 430, loss: 0.00020691368263214827
step: 440, loss: 4.6246797865023836e-05
step: 450, loss: 2.8661688702413812e-05
step: 460, loss: 2.26862703129882e-05
step: 470, loss: 0.0004276178078725934
step: 480, loss: 0.002881467342376709
step: 490, loss: 0.00028200005181133747
step: 500, loss: 3.86815590900369e-05
step: 510, loss: 3.053160253330134e-05
step: 520, loss: 2.5074446966755204e-05
step: 530, loss: 0.00021520693553611636
/usr/local/lib/python3.7/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=0.9302107728337237, f1=0.9240986717267552,
best_f1=0.9269195189639223
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00037034929846413434
step: 10, loss: 0.00018993655976373702
step: 20, loss: 4.197578164166771e-05
step: 30, loss: 2.879586463677697e-05
step: 40, loss: 2.1967471184325404e-05
step: 50, loss: 8.205498306779191e-05
step: 60, loss: 3.567561725503765e-05
step: 70, loss: 2.3543032511952333e-05
step: 80, loss: 5.9154081100132316e-05
step: 90, loss: 1.7832720914157107e-05
step: 100, loss: 6.428598135244101e-05
step: 110, loss: 2.9066910428809933e-05
step: 120, loss: 3.4393087844364345e-05
step: 130, loss: 4.752684981212951e-05
step: 140, loss: 0.00010346341878175735
step: 150, loss: 0.00013034399307798594
step: 160, loss: 0.00013615342322736979
step: 170, loss: 0.00015852137585170567
step: 180, loss: 3.164857116644271e-05
step: 190, loss: 3.2572937925579026e-05
step: 200, loss: 5.471720214700326e-05
step: 210, loss: 5.525867163669318e-05
step: 220, loss: 4.230805279803462e-05

```

```

step: 230, loss: 0.003619135357439518
step: 240, loss: 3.0210689146770164e-05
step: 250, loss: 3.968277451349422e-05
step: 260, loss: 2.0403054804774e-05
step: 270, loss: 0.0018039030255749822
step: 280, loss: 3.363944051670842e-05
step: 290, loss: 0.0001148948649642989
step: 300, loss: 0.00022847697255201638
step: 310, loss: 2.60202959907474e-05
step: 320, loss: 0.00032911068410612643
step: 330, loss: 9.587071690475568e-05
step: 340, loss: 3.6979174183215946e-05
step: 350, loss: 0.012959806248545647
step: 360, loss: 0.0003171641146764159
step: 370, loss: 4.230369449942373e-05
step: 380, loss: 2.4176313672796823e-05
step: 390, loss: 2.4325629055965692e-05
step: 400, loss: 2.8378468414302915e-05
step: 410, loss: 2.5908728275680915e-05
step: 420, loss: 3.648627898655832e-05
step: 430, loss: 3.21368315781001e-05
step: 440, loss: 1.7072772607207298e-05
step: 450, loss: 2.1978767108521424e-05
step: 460, loss: 5.771701762569137e-05
step: 470, loss: 3.562538040569052e-05
step: 480, loss: 2.4433506041532382e-05
step: 490, loss: 0.00256637972779572
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 500, loss: 0.00014722722698934376
step: 510, loss: 7.425771764246747e-05
step: 520, loss: 5.9919933846686035e-05
step: 530, loss: 3.0084462196100503e-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=0.9266917293233082, f1=0.9252601702932829,
best_f1=0.9269195189639223
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.491203046403825e-05
step: 10, loss: 0.00014672330871690065
step: 20, loss: 2.3308708477998152e-05
step: 30, loss: 4.057304249727167e-05
step: 40, loss: 4.036883910885081e-05
step: 50, loss: 9.636867616791278e-05
step: 60, loss: 2.2593272660742514e-05
step: 70, loss: 5.2548039093380794e-05

```


step: 80, loss: 5.086988312541507e-05
step: 90, loss: 3.099171954090707e-05
step: 100, loss: 2.6050236556329764e-05
step: 110, loss: 0.00030630521359853446
step: 120, loss: 0.001669945428147912
step: 130, loss: 6.945112545508891e-05
step: 140, loss: 2.3349288312601857e-05
step: 150, loss: 6.198129267431796e-05
step: 160, loss: 3.048217149625998e-05
step: 170, loss: 4.1459778003627434e-05
step: 180, loss: 2.2299165721051395e-05
step: 190, loss: 0.00013471879356075078
step: 200, loss: 3.701184323290363e-05
step: 210, loss: 3.6668101529357955e-05
step: 220, loss: 2.0067771401954815e-05
step: 230, loss: 4.802383773494512e-05
step: 240, loss: 2.8840517188655213e-05
step: 250, loss: 3.330681283841841e-05
step: 260, loss: 4.3130956328241155e-05
step: 270, loss: 6.717950600432232e-05
step: 280, loss: 6.517394649563357e-05
step: 290, loss: 0.00039060102426446974
step: 300, loss: 0.00047363474732264876
step: 310, loss: 2.7108211725135334e-05
step: 320, loss: 3.286348874098621e-05
step: 330, loss: 0.0003010074724443257
step: 340, loss: 9.691625746199861e-05
step: 350, loss: 3.799258411163464e-05
step: 360, loss: 8.362979133380577e-05
step: 370, loss: 0.00012834920198656619
step: 380, loss: 2.6750642064143904e-05
step: 390, loss: 2.5085260858759284e-05
step: 400, loss: 3.850966095342301e-05
step: 410, loss: 4.755134796141647e-05
step: 420, loss: 0.0001325467455899343
step: 430, loss: 2.378554199822247e-05
step: 440, loss: 0.0002384374529356137
step: 450, loss: 0.0001361922186333686
step: 460, loss: 2.347975350858178e-05
step: 470, loss: 4.446405728231184e-05
step: 480, loss: 0.0003816541866399348
step: 490, loss: 7.846275548217818e-05
step: 500, loss: 4.552110840450041e-05
step: 510, loss: 5.7259818277088925e-05
step: 520, loss: 3.5413297155173495e-05
step: 530, loss: 0.0001261662837350741
/usr/local/lib/python3.7/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=0.9284064665127021, f1=0.9258572752548655,
best_f1=0.9269195189639223
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
5742it [00:17, 320.00it/s]
load_f1 = 0.9299129638112689
real_f1 = 0.9305301645338208
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, 352.27it/s]

```

3 DK

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

Looking in indexes: <https://pypi.org/simple>, <https://us-python.pkg.dev/colab-wheels/public/simple/>

Collecting en_core_web_lg==2.2.5

Downloading https://github.com/explosion/spacy-models/releases/download/en_core_web_lg-2.2.5/en_core_web_lg-2.2.5.tar.gz (827.9 MB)

| | 827.9 MB 1.2 MB/s

Requirement already satisfied: spacy>=2.2.2 in
/usr/local/lib/python3.7/dist-packages (from en_core_web_lg==2.2.5) (2.2.3)

Requirement already satisfied: 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: 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: 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: 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: 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: 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: 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: 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: 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: 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: 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: 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: certifi>=2017.4.17 in
/usr/local/lib/python3.7/dist-packages (from
requests<3.0.0,>=2.13.0->spacy>=2.2.2->en_core_web_lg==2.2.5) (2022.6.15)
Requirement already satisfied: chardet<4,>=3.0.2 in
/usr/local/lib/python3.7/dist-packages (from
requests<3.0.0,>=2.13.0->spacy>=2.2.2->en_core_web_lg==2.2.5) (3.0.4)
Requirement already satisfied: 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: 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: 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=79864db052ee51624f5fdb621bde513b35bd2aacdf55bfc5c1d1daeffeebddab
  Stored in directory: /tmp/pip-ephem-wheel-cache-d1d4q_xd/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)...

```

After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.854921817779541
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=0.2692307692307693, f1=0.2666666666666667,
best_f1=0.2666666666666667
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.37128689885139465
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.4666666666666667, f1=0.358974358974359,
best_f1=0.358974358974359

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.31033003330230713
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.4878048780487805, f1=0.4210526315789474,
best_f1=0.4210526315789474
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3391956686973572
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5238095238095237, f1=0.4363636363636364,
best_f1=0.4363636363636364
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2237047255039215
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6206896551724138, f1=0.43478260869565216,
best_f1=0.43478260869565216
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.22111691534519196
/usr/local/lib/python3.7/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=0.5714285714285714, f1=0.2857142857142857,
best_f1=0.43478260869565216

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2496025711297989
/usr/local/lib/python3.7/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=0.56, f1=0.4444444444444444, best_f1=0.43478260869565216
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2814851701259613
/usr/local/lib/python3.7/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=0.6086956521739131, f1=0.47058823529411764,
best_f1=0.43478260869565216
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1287187933921814
/usr/local/lib/python3.7/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=0.6086956521739131, f1=0.3333333333333333,
best_f1=0.43478260869565216
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1958571970462799
/usr/local/lib/python3.7/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=0.6086956521739131, f1=0.4242424242424242,
best_f1=0.43478260869565216
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2234392911195755
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6666666666666666, f1=0.39999999999999997,
best_f1=0.39999999999999997

```



```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.20809686183929443
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.6666666666666666, f1=0.4, best_f1=0.39999999999999997
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1218341812491417
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 13: dev_f1=0.6666666666666666, f1=0.4242424242424242,
best_f1=0.39999999999999997
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.11183898150920868
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.6666666666666666, f1=0.4242424242424242,
best_f1=0.39999999999999997
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.24612870812416077
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 15: dev_f1=0.6666666666666666, f1=0.4242424242424242,
best_f1=0.39999999999999997
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

```

Defaults for this optimization level are:

```

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

```

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

After processing overrides, optimization options are:

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

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

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

```
91it [00:00, 132528.36it/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 = 0.6666666666666666
```

```
real_f1 = 0.6060606060606061
```

```
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, 381.51it/s]
```

3.1.2 DBLP-ACM - Running the matcher

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

```

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

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

```

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

Defaults for this optimization level are:

```

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

```

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

After processing overrides, optimization options are:

```

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

```

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

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

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

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

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

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

```

"https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8009310960769653
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.4737609028816223
step: 20, loss: 0.610133171081543
step: 30, loss: 0.45605722069740295
step: 40, loss: 0.23413918912410736
step: 50, loss: 0.09919269382953644
step: 60, loss: 0.08076944947242737
step: 70, loss: 0.24102789163589478
step: 80, loss: 0.17148414254188538
step: 90, loss: 0.03574023395776749
step: 100, loss: 0.1108587235212326
step: 110, loss: 0.07693708688020706
step: 120, loss: 0.06370244920253754
step: 130, loss: 0.011808207258582115
step: 140, loss: 0.01285028737038374
step: 150, loss: 0.07954387366771698
step: 160, loss: 0.11761096119880676
step: 170, loss: 0.008767654187977314
step: 180, loss: 0.013487865217030048
step: 190, loss: 0.007891225628554821
step: 200, loss: 0.003887270577251911
step: 210, loss: 0.0025701746344566345
step: 220, loss: 0.0033152811229228973
step: 230, loss: 0.10129514336585999
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9683972911963882, f1=0.9597238204833141,
best_f1=0.9597238204833141
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010113474912941456
step: 10, loss: 0.004210702609270811
step: 20, loss: 0.013602757826447487
step: 30, loss: 0.003498131176456809
step: 40, loss: 0.0896526500582695
step: 50, loss: 0.005641581024974585
step: 60, loss: 0.0029467621352523565
step: 70, loss: 0.004637897480279207
step: 80, loss: 0.0025535800959914923

```

```

step: 90, loss: 0.07037501782178879
step: 100, loss: 0.016138646751642227
step: 110, loss: 0.027651289477944374
step: 120, loss: 0.04361969232559204
step: 130, loss: 0.002548663644120097
step: 140, loss: 0.03245360031723976
step: 150, loss: 0.0110047971829772
step: 160, loss: 0.027489013969898224
step: 170, loss: 0.005884531885385513
step: 180, loss: 0.00808760616928339
step: 190, loss: 0.03591453656554222
step: 200, loss: 0.0031313535291701555
step: 210, loss: 0.0531449019908905
step: 220, loss: 0.00313437613658607
step: 230, loss: 0.009167516604065895
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9729119638826186, f1=0.9622857142857143,
best_f1=0.9622857142857143
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.025286512449383736
step: 10, loss: 0.003466837340965867
step: 20, loss: 0.005566410254687071
step: 30, loss: 0.011021408252418041
step: 40, loss: 0.02988257445394993
step: 50, loss: 0.0073632472194731236
step: 60, loss: 0.003259300021454692
step: 70, loss: 0.000702348945196718
step: 80, loss: 0.0061334907077252865
step: 90, loss: 0.0005824292311444879
step: 100, loss: 0.0009059036383405328
step: 110, loss: 0.0019370692316442728
step: 120, loss: 0.0012405741726979613
step: 130, loss: 0.00041741743916645646
step: 140, loss: 0.014574370346963406
step: 150, loss: 0.05108753591775894
step: 160, loss: 0.005093414802104235
step: 170, loss: 0.007741441018879414
step: 180, loss: 0.00336312223225832
step: 190, loss: 0.012029288336634636
step: 200, loss: 0.0015427196631208062

```

```

step: 210, loss: 0.018532713875174522
step: 220, loss: 0.0018166849622502923
step: 230, loss: 0.14732815325260162
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.984304932735426, f1=0.9798657718120806,
best_f1=0.9798657718120806
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008639393490739167
step: 10, loss: 0.042111482471227646
step: 20, loss: 0.0021006599999964237
step: 30, loss: 0.000551806588191539
step: 40, loss: 0.0019009487004950643
step: 50, loss: 0.003978288732469082
step: 60, loss: 0.0015922782476991415
step: 70, loss: 0.006868666037917137
step: 80, loss: 0.07555527985095978
step: 90, loss: 0.004005255177617073
step: 100, loss: 0.00911684986203909
step: 110, loss: 0.0039855437353253365
step: 120, loss: 0.004114147275686264
step: 130, loss: 0.06263694912195206
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 140, loss: 0.002957284450531006
step: 150, loss: 0.0008297623717226088
step: 160, loss: 0.18346400558948517
step: 170, loss: 0.008941487409174442
step: 180, loss: 0.10975220054388046
step: 190, loss: 0.011153988540172577
step: 200, loss: 0.0014053110498934984
step: 210, loss: 0.0010161283425986767
step: 220, loss: 0.005040450487285852
step: 230, loss: 0.0008556543616577983
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9853438556933484, f1=0.9807474518686297,

```

```

best_f1=0.9807474518686297
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008794975583441556
step: 10, loss: 0.0012429705820977688
step: 20, loss: 0.0003867573686875403
step: 30, loss: 0.00023145075829233974
step: 40, loss: 0.0012558528687804937
step: 50, loss: 0.0008600362343713641
step: 60, loss: 0.0003714033227879554
step: 70, loss: 0.00031337293330579996
step: 80, loss: 0.0013627829030156136
step: 90, loss: 0.013532688841223717
step: 100, loss: 0.0050643570721149445
step: 110, loss: 0.0025082214269787073
step: 120, loss: 0.044180549681186676
step: 130, loss: 0.016358882188796997
step: 140, loss: 0.00027605053037405014
step: 150, loss: 0.0002107262407662347
step: 160, loss: 0.09608027338981628
step: 170, loss: 0.043893635272979736
step: 180, loss: 0.0008749460685066879
step: 190, loss: 0.00037613214226439595
step: 200, loss: 0.002375638112425804
step: 210, loss: 0.0003776034864131361
step: 220, loss: 0.0006420793361030519
step: 230, loss: 0.0011802188819274306
/usr/local/lib/python3.7/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=0.9830890642615557, f1=0.9773242630385486,
best_f1=0.9807474518686297
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0012213574955239892
step: 10, loss: 0.00046368330367840827
step: 20, loss: 0.0010570784797891974
step: 30, loss: 0.0009544100612401962
step: 40, loss: 0.0014167556073516607
step: 50, loss: 0.002526214113458991
step: 60, loss: 0.013963642530143261
step: 70, loss: 0.0009180405759252608
step: 80, loss: 0.0005935564986430109
step: 90, loss: 0.000323703745380044
step: 100, loss: 0.000425395934144035
step: 110, loss: 0.05644169822335243

```

```

step: 120, loss: 0.003246621461585164
step: 130, loss: 0.014518054202198982
step: 140, loss: 0.0028485727962106466
step: 150, loss: 0.00046789279440417886
step: 160, loss: 0.0003111388941761106
step: 170, loss: 0.00032908827415667474
step: 180, loss: 0.0012980824103578925
step: 190, loss: 0.001947835786268115
step: 200, loss: 0.0006761755794286728
step: 210, loss: 0.0008566597825847566
step: 220, loss: 0.00031295479857362807
step: 230, loss: 0.0015169023536145687
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9875706214689265, f1=0.9796839729119639,
best_f1=0.9796839729119639
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.16433361172676086
step: 10, loss: 0.0001279843127122149
step: 20, loss: 0.0006939386948943138
step: 30, loss: 0.00017075406503863633
step: 40, loss: 0.0032008017878979445
step: 50, loss: 0.0001335124543402344
step: 60, loss: 0.0006569609395228326
step: 70, loss: 0.0005209152586758137
step: 80, loss: 0.00013043991930317134
step: 90, loss: 0.0003913686377927661
step: 100, loss: 0.013492453843355179
step: 110, loss: 0.00045356189366430044
step: 120, loss: 0.00019907030218746513
step: 130, loss: 0.0008198308642022312
step: 140, loss: 0.00017472686886321753
step: 150, loss: 0.0010249497136101127
step: 160, loss: 0.0002759322233032435
step: 170, loss: 0.00014271527470555156
step: 180, loss: 0.0003021048032678664
step: 190, loss: 0.00022081786300987005
step: 200, loss: 0.02536456100642681
step: 210, loss: 0.0001329516526311636
step: 220, loss: 0.0016096364706754684
step: 230, loss: 0.022036954760551453

```



```

/usr/local/lib/python3.7/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=0.9841986455981941, f1=0.9786276715410572,
best_f1=0.9796839729119639
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02999911829829216
step: 10, loss: 0.029980158433318138
step: 20, loss: 0.00022288276522886008
step: 30, loss: 0.00018465980247128755
step: 40, loss: 0.00010375728743383661
step: 50, loss: 0.0008404079126194119
step: 60, loss: 0.0012739936355501413
step: 70, loss: 0.00014184041356202215
step: 80, loss: 0.00019941911159548908
step: 90, loss: 0.014863613061606884
step: 100, loss: 0.00012507295468822122
step: 110, loss: 0.0008320249617099762
step: 120, loss: 0.00038315620622597635
step: 130, loss: 0.00047583613195456564
step: 140, loss: 0.0010224867146462202
step: 150, loss: 0.006497661583125591
step: 160, loss: 0.00016818029689602554
step: 170, loss: 0.0014470627065747976
step: 180, loss: 0.0055887335911393166
step: 190, loss: 0.0002193148829974234
step: 200, loss: 0.000335051619913429
step: 210, loss: 0.0001598983071744442
step: 220, loss: 0.00017156456306111068
step: 230, loss: 0.05676186457276344
/usr/local/lib/python3.7/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=0.9820627802690582, f1=0.9786276715410572,
best_f1=0.9796839729119639
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00033865647856146097
step: 10, loss: 0.0003074730921071023
step: 20, loss: 0.001980644417926669
step: 30, loss: 0.05428438261151314
step: 40, loss: 0.00010282438597641885
step: 50, loss: 0.0001358051667921245
step: 60, loss: 0.00025706615997478366
step: 70, loss: 0.0005838068900629878

```

```

step: 80, loss: 0.051220331341028214
step: 90, loss: 0.0028459765017032623
step: 100, loss: 0.00014913527411408722
step: 110, loss: 0.00014537143579218537
step: 120, loss: 0.03439709171652794
step: 130, loss: 0.00022322291624732316
step: 140, loss: 0.00016886148659978062
step: 150, loss: 7.701607682975009e-05
step: 160, loss: 0.00013961148215457797
step: 170, loss: 9.52877671807073e-05
step: 180, loss: 0.0010126486886292696
step: 190, loss: 9.198466432280838e-05
step: 200, loss: 0.02462209016084671
step: 210, loss: 0.00010135447519132867
step: 220, loss: 0.030770180746912956
step: 230, loss: 0.0021319680381566286
/usr/local/lib/python3.7/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=0.980963045912654, f1=0.9741863075196409,
best_f1=0.9796839729119639
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00013100040087010711
step: 10, loss: 0.00010433929855935276
step: 20, loss: 0.00010555985500104725
step: 30, loss: 0.0001311855885433033
step: 40, loss: 0.000697125680744648
step: 50, loss: 0.05261337757110596
step: 60, loss: 0.09804032742977142
step: 70, loss: 0.0003439761931076646
step: 80, loss: 0.0009137447341345251
step: 90, loss: 0.00035213763476349413
step: 100, loss: 0.00024324945115949959
step: 110, loss: 0.08414815366268158
step: 120, loss: 0.017169129103422165
step: 130, loss: 0.0002808451245073229
step: 140, loss: 0.08941082656383514
step: 150, loss: 0.001513708382844925
step: 160, loss: 0.0006315078935585916
step: 170, loss: 0.00042504770681262016
step: 180, loss: 0.0001768583315424621
step: 190, loss: 0.00028600444784387946
step: 200, loss: 0.00010301414295099676
step: 210, loss: 7.090051076374948e-05
step: 220, loss: 0.0013289520284160972
step: 230, loss: 0.00014282265328802168

```

```

/usr/local/lib/python3.7/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=0.9864253393665158, f1=0.9830890642615557,
best_f1=0.9796839729119639
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.683272219263017e-05
step: 10, loss: 8.165555482264608e-05
step: 20, loss: 0.0008229466038756073
step: 30, loss: 0.00017695201677270234
step: 40, loss: 0.0002486618177499622
step: 50, loss: 0.00025090473354794085
step: 60, loss: 0.00011490038014017045
step: 70, loss: 0.0003274460032116622
step: 80, loss: 0.19054058194160461
step: 90, loss: 0.02240322344005108
step: 100, loss: 0.000295483652735129
step: 110, loss: 0.00017335043230559677
step: 120, loss: 0.00019074862939305604
step: 130, loss: 8.702781633473933e-05
step: 140, loss: 9.684453107183799e-05
step: 150, loss: 8.342913497472182e-05
step: 160, loss: 0.0002312977594556287
step: 170, loss: 0.014085211791098118
step: 180, loss: 0.0002973606460727751
step: 190, loss: 0.0005692516569979489
step: 200, loss: 0.00013476551976054907
step: 210, loss: 0.00016332720406353474
step: 220, loss: 0.00013974278408568352
step: 230, loss: 0.02775133028626442
/usr/local/lib/python3.7/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=0.9840546697038726, f1=0.9772209567198178,
best_f1=0.9796839729119639
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00037765142042189837
step: 10, loss: 0.02368175983428955
step: 20, loss: 0.0005234446725808084
step: 30, loss: 0.0004741897573694587
step: 40, loss: 6.986518565099686e-05
step: 50, loss: 7.23148332326673e-05
step: 60, loss: 0.010784520767629147
step: 70, loss: 0.00016502555808983743

```

```

step: 80, loss: 5.023584526497871e-05
step: 90, loss: 5.9866659285034984e-05
step: 100, loss: 0.00021954024850856513
step: 110, loss: 8.239890303229913e-05
step: 120, loss: 8.916107617551461e-05
step: 130, loss: 0.00011707313387887552
step: 140, loss: 7.472617289749905e-05
step: 150, loss: 0.00010632555495249107
step: 160, loss: 0.021093929186463356
step: 170, loss: 7.85731608630158e-05
step: 180, loss: 5.891697946935892e-05
step: 190, loss: 0.00014781735080759972
step: 200, loss: 0.002823641523718834
step: 210, loss: 6.835109525127336e-05
step: 220, loss: 0.026811057701706886
step: 230, loss: 0.00022245173749979585
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 12: dev_f1=0.9875706214689265, f1=0.9773755656108598,
best_f1=0.9796839729119639
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0018618196481838822
step: 10, loss: 0.00026650814106687903
step: 20, loss: 0.00033294889726676047
step: 30, loss: 0.026192396879196167
step: 40, loss: 0.00015339007950387895
step: 50, loss: 8.775918831815943e-05
step: 60, loss: 0.00030177479493431747
step: 70, loss: 7.60258044465445e-05
step: 80, loss: 7.028612890280783e-05
step: 90, loss: 0.00012223541853018105
step: 100, loss: 0.00011477506632218137
step: 110, loss: 7.437870226567611e-05
step: 120, loss: 7.126583659555763e-05
step: 130, loss: 8.889608579920605e-05
step: 140, loss: 0.0001352768304059282
step: 150, loss: 0.02546778693795204
step: 160, loss: 9.180212509818375e-05
step: 170, loss: 6.801701238146052e-05
step: 180, loss: 0.00021118443692103028
step: 190, loss: 6.350537296384573e-05
step: 200, loss: 9.19243466341868e-05
step: 210, loss: 0.00011669283412629738
step: 220, loss: 3.653248495538719e-05
step: 230, loss: 4.0532795537728816e-05

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
UserWarning: Please also save or load the state of the optimizer when saving or
loading the scheduler.
    warnings.warn(SAVE_STATE_WARNING, UserWarning)
epoch 13: dev_f1=0.987598647125141, f1=0.9785310734463276,
best_f1=0.9785310734463276
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00048391500604338944
step: 10, loss: 3.719421511050314e-05
step: 20, loss: 0.01262710615992546
step: 30, loss: 0.00011099953553639352
step: 40, loss: 4.0684219129616395e-05
step: 50, loss: 8.072034688666463e-05
step: 60, loss: 4.248411278240383e-05
step: 70, loss: 4.9830094212666154e-05
step: 80, loss: 0.00012736099597532302
step: 90, loss: 0.00013462766946759075
step: 100, loss: 0.004423009697347879
step: 110, loss: 0.00010285707685397938
step: 120, loss: 0.00022735839593224227
step: 130, loss: 8.945931040216237e-05
step: 140, loss: 0.00020912446780130267
step: 150, loss: 5.8479745348449796e-05
step: 160, loss: 4.906192407361232e-05
step: 170, loss: 0.00030704864184372127
step: 180, loss: 8.220433664973825e-05
step: 190, loss: 0.00022846987121738493
step: 200, loss: 7.052806904539466e-05
step: 210, loss: 0.006532811094075441
step: 220, loss: 7.20583921065554e-05
step: 230, loss: 4.567677387967706e-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=0.9875706214689265, f1=0.9785794813979707,
best_f1=0.9785310734463276
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 8.306750532938167e-05
step: 10, loss: 8.259362948592752e-05
step: 20, loss: 7.206006557680666e-05
step: 30, loss: 0.000750199833419174

```

```

step: 40, loss: 0.00013213428610470146
step: 50, loss: 7.380689930869266e-05
step: 60, loss: 3.983940405305475e-05
step: 70, loss: 6.127294909674674e-05
step: 80, loss: 0.036240678280591965
step: 90, loss: 3.4926659282064065e-05
step: 100, loss: 5.9016208979301155e-05
step: 110, loss: 0.00012067894567735493
step: 120, loss: 8.651965617900714e-05
step: 130, loss: 7.68515164963901e-05
step: 140, loss: 0.00010101604857482016
step: 150, loss: 0.00010487745748832822
step: 160, loss: 8.141529542626813e-05
step: 170, loss: 3.607789403758943e-05
step: 180, loss: 5.9756242990260944e-05
step: 190, loss: 0.014032049104571342
step: 200, loss: 8.906282164389268e-05
step: 210, loss: 0.024494675919413567
step: 220, loss: 0.00010251319326926023
step: 230, loss: 9.129176032729447e-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=0.987598647125141, f1=0.9797297297297298,
best_f1=0.9785310734463276
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.

```

```

Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
2473it [00:08, 306.26it/s]
load_f1 = 0.9875706214689265
real_f1 = 0.9841986455981941
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:12, 358.53it/s]

```

3.1.3 DBLP-GoogleScholar - Running the matcher

```

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

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

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

```

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

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

Defaults for this optimization level are:

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

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

After processing overrides, optimization options are:

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

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

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

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

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

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

```
step: 0, loss: 0.7936778664588928
```

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

```
step: 10, loss: 0.4538373053073883
```

```
step: 20, loss: 0.49497246742248535
```

```
step: 30, loss: 0.40101194381713867
```

```
step: 40, loss: 0.2883652448654175
```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0

```
step: 50, loss: 0.49295574426651
```


step: 60, loss: 0.4714761972427368
step: 70, loss: 0.09675777703523636
step: 80, loss: 0.11442537605762482
step: 90, loss: 0.14720584452152252
step: 100, loss: 0.2662293016910553
step: 110, loss: 0.14127105474472046
step: 120, loss: 0.07311360538005829
step: 130, loss: 0.034435175359249115
step: 140, loss: 0.32998767495155334
step: 150, loss: 0.04317164048552513
step: 160, loss: 0.24861125648021698
step: 170, loss: 0.3093024492263794
step: 180, loss: 0.09928981214761734
step: 190, loss: 0.0780310109257698
step: 200, loss: 0.10273650288581848
step: 210, loss: 0.12289604544639587
step: 220, loss: 0.1439925581216812
step: 230, loss: 0.1622825711965561
step: 240, loss: 0.03634463995695114
step: 250, loss: 0.08505880832672119
step: 260, loss: 0.013607042841613293
step: 270, loss: 0.04035701975226402
step: 280, loss: 0.15082208812236786
step: 290, loss: 0.0664641261100769
step: 300, loss: 0.09796135127544403
step: 310, loss: 0.0705217570066452
step: 320, loss: 0.04908401519060135
step: 330, loss: 0.21435914933681488
step: 340, loss: 0.15923498570919037
step: 350, loss: 0.10890669375658035
step: 360, loss: 0.1000739336013794
step: 370, loss: 0.1676630824804306
step: 380, loss: 0.1255660206079483
step: 390, loss: 0.034015100449323654
step: 400, loss: 0.008692660368978977
step: 410, loss: 0.017557984218001366
step: 420, loss: 0.031535953283309937
step: 430, loss: 0.07846292108297348
step: 440, loss: 0.16592815518379211
step: 450, loss: 0.049194250255823135
step: 460, loss: 0.057326823472976685
step: 470, loss: 0.16349506378173828
step: 480, loss: 0.2825734317302704
step: 490, loss: 0.11694226413965225
step: 500, loss: 0.024504009634256363
step: 510, loss: 0.16373330354690552
step: 520, loss: 0.10796218365430832
step: 530, loss: 0.15617002546787262

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9125057260650481, f1=0.9215328467153284,
best_f1=0.9215328467153284
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10368062555789948
step: 10, loss: 0.1428728997707367
step: 20, loss: 0.1420748084783554
step: 30, loss: 0.04468698799610138
step: 40, loss: 0.03503335639834404
step: 50, loss: 0.18416759371757507
step: 60, loss: 0.2142535150051117
step: 70, loss: 0.1541278064250946
step: 80, loss: 0.04498237371444702
step: 90, loss: 0.0728549212217331
step: 100, loss: 0.205058291554451
step: 110, loss: 0.02598031610250473
step: 120, loss: 0.05589849874377251
step: 130, loss: 0.024705136194825172
step: 140, loss: 0.03512131795287132
step: 150, loss: 0.07121209800243378
step: 160, loss: 0.01229789387434721
step: 170, loss: 0.1009795293211937
step: 180, loss: 0.030750827863812447
step: 190, loss: 0.009314342401921749
step: 200, loss: 0.014922828413546085
step: 210, loss: 0.033273257315158844
step: 220, loss: 0.15895862877368927
step: 230, loss: 0.009060027077794075
step: 240, loss: 0.10765132308006287
step: 250, loss: 0.07689511030912399
step: 260, loss: 0.05915862321853638
step: 270, loss: 0.13422691822052002
step: 280, loss: 0.10023046284914017
step: 290, loss: 0.10583681613206863
step: 300, loss: 0.04602804407477379
step: 310, loss: 0.07408740371465683
step: 320, loss: 0.05553175136446953
step: 330, loss: 0.053480491042137146
step: 340, loss: 0.009013903327286243
step: 350, loss: 0.043110620230436325

```

```

step: 360, loss: 0.0807887390255928
step: 370, loss: 0.008647290989756584
step: 380, loss: 0.09542009979486465
step: 390, loss: 0.035488061606884
step: 400, loss: 0.054552145302295685
step: 410, loss: 0.0022842385806143284
step: 420, loss: 0.11023399233818054
step: 430, loss: 0.034477222710847855
step: 440, loss: 0.04043562710285187
step: 450, loss: 0.012017445638775826
step: 460, loss: 0.2366611361503601
step: 470, loss: 0.07379365712404251
step: 480, loss: 0.24334721267223358
step: 490, loss: 0.06743371486663818
step: 500, loss: 0.021772881969809532
step: 510, loss: 0.06045966222882271
step: 520, loss: 0.04351482540369034
step: 530, loss: 0.1907854676246643
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.922863741339492, f1=0.9279112754158966,
best_f1=0.9279112754158966
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00423922436311841
step: 10, loss: 0.16200877726078033
step: 20, loss: 0.19082118570804596
step: 30, loss: 0.1873035877943039
step: 40, loss: 0.01564623974263668
step: 50, loss: 0.009759831242263317
step: 60, loss: 0.004878113511949778
step: 70, loss: 0.054290786385536194
step: 80, loss: 0.01662403531372547
step: 90, loss: 0.1294449120759964
step: 100, loss: 0.01541794091463089
step: 110, loss: 0.04168218746781349
step: 120, loss: 0.11182013899087906
step: 130, loss: 0.041431739926338196
step: 140, loss: 0.02539115585386753
step: 150, loss: 0.014176126569509506
step: 160, loss: 0.0030440324917435646
step: 170, loss: 0.010725491680204868

```

```

step: 180, loss: 0.009071649052202702
step: 190, loss: 0.006948123686015606
step: 200, loss: 0.05722453445196152
step: 210, loss: 0.029358362779021263
step: 220, loss: 0.034297455102205276
step: 230, loss: 0.06055500730872154
step: 240, loss: 0.0024492384400218725
step: 250, loss: 0.06728927046060562
step: 260, loss: 0.022055497393012047
step: 270, loss: 0.0034941595513373613
step: 280, loss: 0.01788618229329586
step: 290, loss: 0.04859679192304611
step: 300, loss: 0.12939505279064178
step: 310, loss: 0.14811760187149048
step: 320, loss: 0.1487932950258255
step: 330, loss: 0.0065416451543569565
step: 340, loss: 0.0036344537511467934
step: 350, loss: 0.014093715697526932
step: 360, loss: 0.026055730879306793
step: 370, loss: 0.00784227903932333
step: 380, loss: 0.016646364703774452
step: 390, loss: 0.01701541431248188
step: 400, loss: 0.023216335102915764
step: 410, loss: 0.01055427361279726
step: 420, loss: 0.03944822773337364
step: 430, loss: 0.026545675471425056
step: 440, loss: 0.11442790180444717
step: 450, loss: 0.08856166154146194
step: 460, loss: 0.12028538435697556
step: 470, loss: 0.018115492537617683
step: 480, loss: 0.03206261247396469
step: 490, loss: 0.011617619544267654
step: 500, loss: 0.08402056246995926
step: 510, loss: 0.0038488972932100296
step: 520, loss: 0.007002358324825764
step: 530, loss: 0.014330685138702393
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9310508796956729, f1=0.9230769230769231,
best_f1=0.9230769230769231
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")

```

step: 0, loss: 0.0022305664606392384
step: 10, loss: 0.01290091685950756
step: 20, loss: 0.029129300266504288
step: 30, loss: 0.005422776099294424
step: 40, loss: 0.0061338418163359165
step: 50, loss: 0.012624818831682205
step: 60, loss: 0.002772017614915967
step: 70, loss: 0.014747205190360546
step: 80, loss: 0.11978255212306976
step: 90, loss: 0.09583025425672531
step: 100, loss: 0.043996457010507584
step: 110, loss: 0.0011696842266246676
step: 120, loss: 0.000651066773571074
step: 130, loss: 0.003457588842138648
step: 140, loss: 0.03321647644042969
step: 150, loss: 0.0006368906470015645
step: 160, loss: 0.0014613195089623332
step: 170, loss: 0.010260936804115772
step: 180, loss: 0.006904180161654949
step: 190, loss: 0.00506730517372489
step: 200, loss: 0.011863810941576958
step: 210, loss: 0.049163248389959335
step: 220, loss: 0.07182851433753967
step: 230, loss: 0.15128080546855927
step: 240, loss: 0.003437632927671075
step: 250, loss: 0.007686927914619446
step: 260, loss: 0.12250524014234543
step: 270, loss: 0.020855983719229698
step: 280, loss: 0.009426555596292019
step: 290, loss: 0.10678835213184357
step: 300, loss: 0.007908222265541553
step: 310, loss: 0.01818176358938217
step: 320, loss: 0.004875387065112591
step: 330, loss: 0.1444082260131836
step: 340, loss: 0.029361527413129807
step: 350, loss: 0.0005232046823948622
step: 360, loss: 0.0016541284276172519
step: 370, loss: 0.0064995973370969296
step: 380, loss: 0.0024615672882646322
step: 390, loss: 0.04027813300490379
step: 400, loss: 0.10340207070112228
step: 410, loss: 0.0800912007689476
step: 420, loss: 0.007983038201928139
step: 430, loss: 0.03724544495344162
step: 440, loss: 0.0738568976521492
step: 450, loss: 0.02057514153420925
step: 460, loss: 0.0145847387611866
step: 470, loss: 0.0045356894843280315

```

step: 480, loss: 0.02217467688024044
step: 490, loss: 0.04511316865682602
step: 500, loss: 0.01598861627280712
step: 510, loss: 0.19497445225715637
step: 520, loss: 0.04405994340777397
step: 530, loss: 0.021864740177989006
/usr/local/lib/python3.7/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=0.927170868347339, f1=0.9176361098185202,
best_f1=0.9230769230769231
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0039455643855035305
step: 10, loss: 0.027324290946125984
step: 20, loss: 0.015633966773748398
step: 30, loss: 0.0004486308025661856
step: 40, loss: 0.04837166145443916
step: 50, loss: 0.010256361216306686
step: 60, loss: 0.0011306548258289695
step: 70, loss: 0.0004510839644353837
step: 80, loss: 0.007442144677042961
step: 90, loss: 0.0038985414430499077
step: 100, loss: 0.034327276051044464
step: 110, loss: 0.02172943204641342
step: 120, loss: 0.005361809395253658
step: 130, loss: 0.0028117510955780745
step: 140, loss: 0.004228500183671713
step: 150, loss: 0.0016680773114785552
step: 160, loss: 0.154209166765213
step: 170, loss: 0.005898796953260899
step: 180, loss: 0.0011216034181416035
step: 190, loss: 0.002767123281955719
step: 200, loss: 0.08069079369306564
step: 210, loss: 0.006611906923353672
step: 220, loss: 0.0006341254920698702
step: 230, loss: 0.0018089764053002
step: 240, loss: 0.034188441932201385
step: 250, loss: 0.0010815953137353063
step: 260, loss: 0.0009880096185952425
step: 270, loss: 0.04263751208782196
step: 280, loss: 0.11488951742649078
step: 290, loss: 0.0017033533658832312
step: 300, loss: 0.12043233960866928
step: 310, loss: 0.0007327148923650384
step: 320, loss: 0.007924258708953857
step: 330, loss: 0.014453991316258907

```

```

step: 340, loss: 0.001292656292207539
step: 350, loss: 0.029891829937696457
step: 360, loss: 0.031976476311683655
step: 370, loss: 0.0019057246390730143
step: 380, loss: 0.08646842837333679
step: 390, loss: 0.012086513452231884
step: 400, loss: 0.01360249798744917
step: 410, loss: 0.0010801985627040267
step: 420, loss: 0.006900461856275797
step: 430, loss: 0.0011360434582456946
step: 440, loss: 0.00685708224773407
step: 450, loss: 0.003040451556444168
step: 460, loss: 0.001229735091328621
step: 470, loss: 0.011332421563565731
step: 480, loss: 0.0031922152265906334
step: 490, loss: 0.009848695248365402
step: 500, loss: 0.0329100526869297
step: 510, loss: 0.0023379649501293898
step: 520, loss: 0.0012407099129632115
step: 530, loss: 0.004139429423958063
/usr/local/lib/python3.7/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=0.9233593391463975, f1=0.9223702342673403,
best_f1=0.9230769230769231
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0011256185825914145
step: 10, loss: 0.00030437891837209463
step: 20, loss: 0.0039635407738387585
step: 30, loss: 0.0022571866866201162
step: 40, loss: 0.000900094339158386
step: 50, loss: 0.0025647340808063745
step: 60, loss: 0.0020813338924199343
step: 70, loss: 0.01987023465335369
step: 80, loss: 0.00012156813318142667
step: 90, loss: 0.0019541692454367876
step: 100, loss: 0.19460277259349823
step: 110, loss: 0.01628892868757248
step: 120, loss: 0.0024677198380231857
step: 130, loss: 0.0027337041683495045
step: 140, loss: 0.002493854844942689
step: 150, loss: 0.003395074512809515
step: 160, loss: 0.020284105092287064
step: 170, loss: 0.020241573452949524
step: 180, loss: 0.00011379353236407042
step: 190, loss: 0.000570183910895139

```

```
step: 200, loss: 7.739531429251656e-05
step: 210, loss: 0.001498485216870904
step: 220, loss: 0.002087176777422428
step: 230, loss: 0.0003572300192900002
step: 240, loss: 0.005689652636647224
step: 250, loss: 0.0004827178781852126
step: 260, loss: 0.0007358301663771272
step: 270, loss: 0.048502642661333084
step: 280, loss: 0.0026502879336476326
step: 290, loss: 0.02696523442864418
step: 300, loss: 0.001730206422507763
step: 310, loss: 0.0038416546303778887
step: 320, loss: 0.02424709126353264
step: 330, loss: 0.0025688721798360348
step: 340, loss: 0.08023571968078613
step: 350, loss: 0.10743055492639542
step: 360, loss: 0.00022919022012501955
step: 370, loss: 0.00014872851897962391
step: 380, loss: 0.025599202141165733
step: 390, loss: 0.006257576402276754
step: 400, loss: 0.000441143725765869
step: 410, loss: 0.007702636532485485
step: 420, loss: 0.012986852787435055
step: 430, loss: 0.0020206542685627937
step: 440, loss: 0.007122451439499855
step: 450, loss: 0.0029341462068259716
step: 460, loss: 0.011221232824027538
step: 470, loss: 0.0937291756272316
step: 480, loss: 0.009089132770895958
step: 490, loss: 0.0009833074873313308
step: 500, loss: 0.0026224153116345406
step: 510, loss: 0.00023354042787104845
step: 520, loss: 0.0003874192771036178
step: 530, loss: 0.004001135937869549
/usr/local/lib/python3.7/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=0.9224952741020794, f1=0.9168646080760096,
best_f1=0.9230769230769231
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02073562890291214
step: 10, loss: 0.005349715705960989
step: 20, loss: 0.0011115892557427287
step: 30, loss: 0.002629396738484502
step: 40, loss: 0.00011395640467526391
step: 50, loss: 0.00037836056435480714
```


step: 60, loss: 0.0001433303696103394
step: 70, loss: 0.0390189103782177
step: 80, loss: 0.23639290034770966
step: 90, loss: 0.0003303898556623608
step: 100, loss: 0.0017358183395117521
step: 110, loss: 0.00037724748835898936
step: 120, loss: 0.0001474337768740952
step: 130, loss: 0.00018512271344661713
step: 140, loss: 0.00033601245377212763
step: 150, loss: 0.00011828293645521626
step: 160, loss: 0.0016003923956304789
step: 170, loss: 0.0007401025504805148
step: 180, loss: 0.0023598468396812677
step: 190, loss: 0.00016584932745900005
step: 200, loss: 0.006016302853822708
step: 210, loss: 0.001476497738622129
step: 220, loss: 0.0017618316924199462
step: 230, loss: 0.000266164424829185
step: 240, loss: 0.0032675324473530054
step: 250, loss: 0.002121608005836606
step: 260, loss: 0.005999906454235315
step: 270, loss: 0.00017232053505722433
step: 280, loss: 0.008671190589666367
step: 290, loss: 0.0013046242529526353
step: 300, loss: 0.0001862601493485272
step: 310, loss: 0.0003973415296059102
step: 320, loss: 0.006209807004779577
step: 330, loss: 0.00012386094022076577
step: 340, loss: 0.013360410928726196
step: 350, loss: 0.00012896375847049057
step: 360, loss: 0.0019035732839256525
step: 370, loss: 0.006053552962839603
step: 380, loss: 0.0003661111113615334
step: 390, loss: 0.000136559407110326
step: 400, loss: 0.0010610123863443732
step: 410, loss: 0.009583438746631145
step: 420, loss: 0.00023655356199014932
step: 430, loss: 0.00012410187628120184
step: 440, loss: 0.02275134064257145
step: 450, loss: 0.0015994757413864136
step: 460, loss: 0.002126478822901845
step: 470, loss: 0.04059848561882973
step: 480, loss: 0.0004373985866550356
step: 490, loss: 0.027120430022478104
step: 500, loss: 0.003586421487852931
step: 510, loss: 0.00021612287673633546
step: 520, loss: 0.00011075520160375163
step: 530, loss: 0.0003593174915295094

```
/usr/local/lib/python3.7/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=0.9150507848568791, f1=0.9183202584217812,
best_f1=0.9230769230769231
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00115027348510921
step: 10, loss: 0.005984090268611908
step: 20, loss: 0.00034258264349773526
step: 30, loss: 0.0005241265753284097
step: 40, loss: 0.0005014886846765876
step: 50, loss: 0.0003819862613454461
step: 60, loss: 0.00038956303615123034
step: 70, loss: 0.008406450971961021
step: 80, loss: 0.000594868091866374
step: 90, loss: 6.696177297271788e-05
step: 100, loss: 0.00016285674064420164
step: 110, loss: 0.0004488211125135422
step: 120, loss: 0.051705725491046906
step: 130, loss: 0.0005413888720795512
step: 140, loss: 3.9694081351626664e-05
step: 150, loss: 0.0006643224623985589
step: 160, loss: 7.274383824551478e-05
step: 170, loss: 0.00014694691344629973
step: 180, loss: 0.00033356351195834577
step: 190, loss: 0.000612253847066313
step: 200, loss: 0.0021051119547337294
step: 210, loss: 0.0006186877726577222
step: 220, loss: 0.0012102730106562376
step: 230, loss: 0.0068437703885138035
step: 240, loss: 0.00033697797334752977
step: 250, loss: 0.002253826940432191
step: 260, loss: 0.10595884919166565
step: 270, loss: 5.204445915296674e-05
step: 280, loss: 0.022821128368377686
step: 290, loss: 0.004264862276613712
step: 300, loss: 0.0014566510217264295
step: 310, loss: 0.016050532460212708
step: 320, loss: 0.002701361896470189
step: 330, loss: 0.018381111323833466
step: 340, loss: 0.0004847378295380622
step: 350, loss: 0.0002601641754154116
step: 360, loss: 0.00019602186512202024
step: 370, loss: 0.029089458286762238
step: 380, loss: 0.0005539124831557274
step: 390, loss: 0.0002038456150330603
```

```

step: 400, loss: 0.2621065676212311
step: 410, loss: 0.000523310387507081
step: 420, loss: 0.0004138736112508923
step: 430, loss: 0.00016712048090994358
step: 440, loss: 0.02060673199594021
step: 450, loss: 0.00021804696007166058
step: 460, loss: 0.0008197898278012872
step: 470, loss: 0.0008188765496015549
step: 480, loss: 0.0003003472229465842
step: 490, loss: 0.0001351883402094245
step: 500, loss: 0.0007672167848795652
step: 510, loss: 7.298702257685363e-05
step: 520, loss: 0.00016992189921438694
step: 530, loss: 7.047505641821772e-05
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 8: dev_f1=0.9232889297197979, f1=0.916628281897743,
best_f1=0.9230769230769231
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00013858938473276794
step: 10, loss: 7.641335832886398e-05
step: 20, loss: 0.00013291325012687594
step: 30, loss: 0.00010730043140938506
step: 40, loss: 0.007447457872331142
step: 50, loss: 0.0046547334641218185
step: 60, loss: 0.0012133694253861904
step: 70, loss: 0.09266036003828049
step: 80, loss: 0.0003573966969270259
step: 90, loss: 7.391029794234782e-05
step: 100, loss: 8.69899449753575e-05
step: 110, loss: 0.0026984389405697584
step: 120, loss: 0.00010774520342238247
step: 130, loss: 0.010234519839286804
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 140, loss: 0.00011407358397264034
step: 150, loss: 0.04133964702486992
step: 160, loss: 7.50418912502937e-05
step: 170, loss: 0.00010035067680291831
step: 180, loss: 6.152418063720688e-05
step: 190, loss: 0.00013361840683501214
step: 200, loss: 0.0005117875407449901
step: 210, loss: 5.325457459548488e-05
step: 220, loss: 0.0003736208309419453
step: 230, loss: 0.00010263996955472976
step: 240, loss: 0.00038193518412299454

```

```

step: 250, loss: 0.0001442155771655962
step: 260, loss: 0.0016878717578947544
step: 270, loss: 0.00021035711688455194
step: 280, loss: 6.343761924654245e-05
step: 290, loss: 5.319164847605862e-05
step: 300, loss: 0.001350211095996201
step: 310, loss: 5.1986131438752636e-05
step: 320, loss: 0.00013787740317638963
step: 330, loss: 0.0003058115253224969
step: 340, loss: 0.000105813451227732
step: 350, loss: 6.849717465229332e-05
step: 360, loss: 0.010070407763123512
step: 370, loss: 0.00016570891602896154
step: 380, loss: 7.37005248083733e-05
step: 390, loss: 0.00018848809122573584
step: 400, loss: 0.012258228845894337
step: 410, loss: 0.00041015350143425167
step: 420, loss: 7.225400622701272e-05
step: 430, loss: 0.00017581068095751107
step: 440, loss: 9.05092092580162e-05
step: 450, loss: 5.724645598093048e-05
step: 460, loss: 9.567602683091536e-05
step: 470, loss: 0.0005326996906660497
step: 480, loss: 0.0007280465797521174
step: 490, loss: 9.538466110825539e-05
step: 500, loss: 0.0005885731661692262
step: 510, loss: 0.00013519175990950316
step: 520, loss: 0.0007152371108531952
step: 530, loss: 0.0008195857517421246
/usr/local/lib/python3.7/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=0.9205607476635514, f1=0.9179869524697111,
best_f1=0.9230769230769231
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007367903599515557
step: 10, loss: 0.0001307086495216936
step: 20, loss: 0.003432017983868718
step: 30, loss: 0.00027094854158349335
step: 40, loss: 0.0022840704768896103
step: 50, loss: 8.204934420064092e-05
step: 60, loss: 0.000154417721205391
step: 70, loss: 0.00014841981464996934
step: 80, loss: 0.0023711153771728277
step: 90, loss: 0.00010728004417615011
step: 100, loss: 0.00016196871001739055

```

```
step: 110, loss: 0.0002085523447021842
step: 120, loss: 0.0014853818574920297
step: 130, loss: 0.00033031057682819664
step: 140, loss: 0.0004729984502773732
step: 150, loss: 7.300001016119495e-05
step: 160, loss: 0.0027650247793644667
step: 170, loss: 0.00020073979976586998
step: 180, loss: 0.02567276544868946
step: 190, loss: 8.140386489685625e-05
step: 200, loss: 0.003033921355381608
step: 210, loss: 0.0001795636781025678
step: 220, loss: 0.00021013329387642443
step: 230, loss: 0.00026829406851902604
step: 240, loss: 0.00012767028238158673
step: 250, loss: 7.023684884188697e-05
step: 260, loss: 0.0002208467631135136
step: 270, loss: 3.7220397643977776e-05
step: 280, loss: 6.57535347272642e-05
step: 290, loss: 0.007197569124400616
step: 300, loss: 0.00038343187770806253
step: 310, loss: 0.0001478760823374614
step: 320, loss: 4.888627154286951e-05
step: 330, loss: 0.0005180825828574598
step: 340, loss: 7.163259579101577e-05
step: 350, loss: 6.090856913942844e-05
step: 360, loss: 9.018374112201855e-05
step: 370, loss: 0.004053322598338127
step: 380, loss: 0.00013491298886947334
step: 390, loss: 8.317029278259724e-05
step: 400, loss: 0.00059803097974509
step: 410, loss: 0.0003491246607154608
step: 420, loss: 0.0005741675267927349
step: 430, loss: 0.00011653600085992366
step: 440, loss: 9.048426727531478e-05
step: 450, loss: 0.0025540657807141542
step: 460, loss: 0.00011330038250889629
step: 470, loss: 7.964269752847031e-05
step: 480, loss: 5.920461262576282e-05
step: 490, loss: 0.07924602925777435
step: 500, loss: 0.0001150467578554526
step: 510, loss: 0.0005400329828262329
step: 520, loss: 0.00039460265543311834
step: 530, loss: 0.0004933182499371469
/usr/local/lib/python3.7/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=0.9228611500701264, f1=0.9156853509185116,
best_f1=0.9230769230769231
```

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

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

step: 0, loss: 0.00013991801824886352
step: 10, loss: 0.0019734774250537157
step: 20, loss: 0.00010226680751657113
step: 30, loss: 0.0027239287737756968
step: 40, loss: 0.009111443534493446
step: 50, loss: 0.0005264768842607737
step: 60, loss: 0.00015117228031158447
step: 70, loss: 8.647199138067663e-05
step: 80, loss: 0.0009340401738882065
step: 90, loss: 0.00015068662469275296
step: 100, loss: 0.00023432580928783864
step: 110, loss: 0.00023870563018135726
step: 120, loss: 0.00024173471319954842
step: 130, loss: 0.017030367627739906
step: 140, loss: 7.346237543970346e-05
step: 150, loss: 0.00013081984070595354
step: 160, loss: 0.00031662528635933995
step: 170, loss: 0.022962626069784164
step: 180, loss: 6.561501504620537e-05
step: 190, loss: 0.0001811296388041228
step: 200, loss: 0.007164438255131245
step: 210, loss: 0.00016486160166095942
step: 220, loss: 0.0008588903001509607
step: 230, loss: 4.619068204192445e-05
step: 240, loss: 0.00013940411736257374
step: 250, loss: 0.0013082935474812984
step: 260, loss: 5.4204541811486706e-05
step: 270, loss: 7.258831465151161e-05
step: 280, loss: 8.05477611720562e-05
step: 290, loss: 7.665946759516373e-05
step: 300, loss: 0.04642435163259506
step: 310, loss: 0.00035663560265675187
step: 320, loss: 0.00021909608040004969
step: 330, loss: 0.0008843307732604444
step: 340, loss: 0.00014050088066142052
step: 350, loss: 0.0001711142249405384
step: 360, loss: 0.0003829375491477549
step: 370, loss: 8.250684913946316e-05
step: 380, loss: 6.103983469074592e-05
step: 390, loss: 0.0037111658602952957
step: 400, loss: 5.296404560795054e-05
step: 410, loss: 4.47642887593247e-05
step: 420, loss: 0.00023036087804939598
step: 430, loss: 0.0002913322241511196
step: 440, loss: 0.00045684288488700986

```
step: 450, loss: 0.00018073977844323963
step: 460, loss: 0.0015576491132378578
step: 470, loss: 0.00036386598367244005
step: 480, loss: 0.00011514482321217656
step: 490, loss: 0.0009506245260126889
step: 500, loss: 0.0002071245398838073
step: 510, loss: 7.639463001396507e-05
step: 520, loss: 6.731391476932913e-05
step: 530, loss: 4.929949136567302e-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=0.9190968955785513, f1=0.9165876777251185,
best_f1=0.9230769230769231
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00019412129768170416
step: 10, loss: 0.0003978945896960795
step: 20, loss: 7.529021968366578e-05
step: 30, loss: 0.000334928889060393
step: 40, loss: 0.00025908747920766473
step: 50, loss: 9.80971526587382e-05
step: 60, loss: 0.00010514063615119085
step: 70, loss: 0.00032340711914002895
step: 80, loss: 5.8450274082133546e-05
step: 90, loss: 0.005206052213907242
step: 100, loss: 3.5385179216973484e-05
step: 110, loss: 0.00015984592027962208
step: 120, loss: 0.0001486287801526487
step: 130, loss: 6.644531822530553e-05
step: 140, loss: 6.941316678421572e-05
step: 150, loss: 0.0003922779578715563
step: 160, loss: 8.539400005247444e-05
step: 170, loss: 2.6482348403078504e-05
step: 180, loss: 0.0008127521141432226
step: 190, loss: 9.549412061460316e-05
step: 200, loss: 0.00037961837369948626
step: 210, loss: 0.001076321117579937
step: 220, loss: 4.6685629058629274e-05
step: 230, loss: 6.0008973377989605e-05
step: 240, loss: 0.0004231741768307984
step: 250, loss: 5.633422188111581e-05
step: 260, loss: 0.014380722306668758
step: 270, loss: 0.00013996190682519227
step: 280, loss: 6.352637865347788e-05
step: 290, loss: 0.0014310331316664815
step: 300, loss: 0.0004309766227379441
```

```

step: 310, loss: 0.001105166389606893
step: 320, loss: 0.00022790815273765475
step: 330, loss: 4.583652116707526e-05
step: 340, loss: 0.0006741933175362647
step: 350, loss: 0.00015864869055803865
step: 360, loss: 3.814373121713288e-05
step: 370, loss: 4.8265152145177126e-05
step: 380, loss: 5.746056558564305e-05
step: 390, loss: 9.019839490065351e-05
step: 400, loss: 7.39894894650206e-05
step: 410, loss: 0.00019100804638583213
step: 420, loss: 0.22113730013370514
step: 430, loss: 0.00029859074857085943
step: 440, loss: 0.0022455905564129353
step: 450, loss: 0.00027379131643101573
step: 460, loss: 0.0007722717709839344
step: 470, loss: 0.0004379722522571683
step: 480, loss: 0.00018719614308793098
step: 490, loss: 5.994446109980345e-05
step: 500, loss: 0.010547259822487831
step: 510, loss: 0.000142527642310597
step: 520, loss: 0.0003644553362391889
step: 530, loss: 0.0005346012767404318
/usr/local/lib/python3.7/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=0.9258572752548655, f1=0.9218604651162791,
best_f1=0.9230769230769231
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.57852612878196e-05
step: 10, loss: 5.136946856509894e-05
step: 20, loss: 0.0001396549923811108
step: 30, loss: 0.005493404343724251
step: 40, loss: 5.815580880152993e-05
step: 50, loss: 8.231824904214591e-05
step: 60, loss: 7.527827256126329e-05
step: 70, loss: 5.605840124189854e-05
step: 80, loss: 4.5032942580292e-05
step: 90, loss: 0.005835449323058128
step: 100, loss: 5.8695226471172646e-05
step: 110, loss: 0.00029910358716733754
step: 120, loss: 0.00011564359738258645
step: 130, loss: 0.0001574825873831287
step: 140, loss: 7.121164526324719e-05
step: 150, loss: 7.047042163321748e-05
step: 160, loss: 0.00010543804819462821

```



```

step: 170, loss: 7.808206282788888e-05
step: 180, loss: 5.4004660341888666e-05
step: 190, loss: 0.0005662540788762271
step: 200, loss: 0.0055981045588850975
step: 210, loss: 0.0015965853817760944
step: 220, loss: 7.892263965914026e-05
step: 230, loss: 0.0009845945751294494
step: 240, loss: 8.726718806428835e-05
step: 250, loss: 0.0020695587154477835
step: 260, loss: 4.2923806176986545e-05
step: 270, loss: 0.00010439539619255811
step: 280, loss: 7.11251559550874e-05
step: 290, loss: 5.5188851547427475e-05
step: 300, loss: 9.852933726506308e-05
step: 310, loss: 4.7738049033796415e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 320, loss: 0.0016655509825795889
step: 330, loss: 0.00010193706839345396
step: 340, loss: 7.054483285173774e-05
step: 350, loss: 0.2377966344356537
step: 360, loss: 6.667742854915559e-05
step: 370, loss: 5.500431143445894e-05
step: 380, loss: 0.00037249925662763417
step: 390, loss: 0.00023979859543032944
step: 400, loss: 4.0096583688864484e-05
step: 410, loss: 0.00036136392736807466
step: 420, loss: 7.22137265256606e-05
step: 430, loss: 0.00024312200548592955
step: 440, loss: 7.400319736916572e-05
step: 450, loss: 7.839997124392539e-05
step: 460, loss: 3.776481025852263e-05
step: 470, loss: 0.00016923360817600042
step: 480, loss: 0.0001516344927949831
step: 490, loss: 0.00024610309628769755
step: 500, loss: 8.701116894371808e-05
step: 510, loss: 5.1751965656876564e-05
step: 520, loss: 4.917286423733458e-05
step: 530, loss: 5.958394831395708e-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=0.9251510925151093, f1=0.9207828518173344,
best_f1=0.9230769230769231
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.284425373654813e-05
step: 10, loss: 0.0001052858933690004

```

step: 20, loss: 0.000465230958070606
step: 30, loss: 7.57314483053051e-05
step: 40, loss: 0.00021811325859744102
step: 50, loss: 5.500621773535386e-05
step: 60, loss: 0.002425552112981677
step: 70, loss: 8.48913550726138e-05
step: 80, loss: 0.000238250067923218
step: 90, loss: 0.0003796224482357502
step: 100, loss: 9.542725456412882e-05
step: 110, loss: 5.3364186896942556e-05
step: 120, loss: 2.8980961360502988e-05
step: 130, loss: 0.00031555944588035345
step: 140, loss: 0.0006818248657509685
step: 150, loss: 0.00010539424692979082
step: 160, loss: 0.000363389088306576
step: 170, loss: 6.626086542382836e-05
step: 180, loss: 4.968485882272944e-05
step: 190, loss: 0.0002285842056153342
step: 200, loss: 6.95181151968427e-05
step: 210, loss: 9.338227391708642e-05
step: 220, loss: 5.769148265244439e-05
step: 230, loss: 0.0003308540035504848
step: 240, loss: 4.709654240286909e-05
step: 250, loss: 6.182410288602114e-05
step: 260, loss: 2.7536538254935294e-05
step: 270, loss: 0.000620986451394856
step: 280, loss: 5.3898398618912324e-05
step: 290, loss: 9.774159116204828e-05
step: 300, loss: 5.176767808734439e-05
step: 310, loss: 4.0122649807017297e-05
step: 320, loss: 0.00044456077739596367
step: 330, loss: 5.303544094203971e-05
step: 340, loss: 5.9002861235057935e-05
step: 350, loss: 0.00012800544209312648
step: 360, loss: 0.0018626231467351317
step: 370, loss: 3.9850528992246836e-05
step: 380, loss: 5.569389395532198e-05
step: 390, loss: 4.785117198480293e-05
step: 400, loss: 5.395802509156056e-05
step: 410, loss: 0.00030677314498461783
step: 420, loss: 3.802548235398717e-05
step: 430, loss: 7.866771920816973e-05
step: 440, loss: 2.562559166108258e-05
step: 450, loss: 3.8685240724589676e-05
step: 460, loss: 6.59743309370242e-05
step: 470, loss: 3.602170181693509e-05
step: 480, loss: 4.010088014183566e-05
step: 490, loss: 7.171343168010935e-05

```

step: 500, loss: 0.0002913315547630191
step: 510, loss: 4.1155006329063326e-05
step: 520, loss: 5.305984814185649e-05
step: 530, loss: 4.410138717503287e-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=0.9280074314909429, f1=0.9240093240093239,
best_f1=0.9230769230769231
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.10805954039096832
step: 10, loss: 0.00011210727825528011
step: 20, loss: 4.979209916200489e-05
step: 30, loss: 5.723110371036455e-05
step: 40, loss: 2.7040576242143288e-05
step: 50, loss: 0.0020337689202278852
step: 60, loss: 3.205045868526213e-05
step: 70, loss: 4.9410955398343503e-05
step: 80, loss: 7.299748540390283e-05
step: 90, loss: 0.0020469084847718477
step: 100, loss: 5.492288983077742e-05
step: 110, loss: 0.00022168215946294367
step: 120, loss: 0.00040444082696922123
step: 130, loss: 4.953506140736863e-05
step: 140, loss: 2.7934982426813804e-05
step: 150, loss: 8.661449101055041e-05
step: 160, loss: 5.468187009682879e-05
step: 170, loss: 7.992584869498387e-05
step: 180, loss: 6.134423892945051e-05
step: 190, loss: 0.0001684489252511412
step: 200, loss: 3.608893166529015e-05
step: 210, loss: 5.059776594862342e-05
step: 220, loss: 2.3870978111517616e-05
step: 230, loss: 0.0015080252196639776
step: 240, loss: 8.344659727299586e-05
step: 250, loss: 7.478863699361682e-05
step: 260, loss: 6.908069190103561e-05
step: 270, loss: 0.00011851725139422342
step: 280, loss: 4.617636659531854e-05
step: 290, loss: 0.0006248755380511284
step: 300, loss: 5.196847632760182e-05
step: 310, loss: 0.00015175496810115874
step: 320, loss: 6.0370082792360336e-05
step: 330, loss: 5.960487760603428e-05
step: 340, loss: 6.173526344355196e-05
step: 350, loss: 7.704651216045022e-05

```

```

step: 360, loss: 6.622433284064755e-05
step: 370, loss: 4.206259109196253e-05
step: 380, loss: 5.288919055601582e-05
step: 390, loss: 6.713972834404558e-05
step: 400, loss: 4.392653499962762e-05
step: 410, loss: 0.0016369220102205873
step: 420, loss: 7.635065412614495e-05
step: 430, loss: 3.618248229031451e-05
step: 440, loss: 9.460572618991137e-05
step: 450, loss: 8.224344492191449e-05
step: 460, loss: 5.178435822017491e-05
step: 470, loss: 3.0307561246445403e-05
step: 480, loss: 0.0027680324856191874
step: 490, loss: 5.3844607464270666e-05
step: 500, loss: 6.941599713172764e-05
step: 510, loss: 7.150983583414927e-05
step: 520, loss: 5.981987851555459e-05
step: 530, loss: 5.2064751798752695e-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=0.924860853432282, f1=0.9215686274509803,
best_f1=0.9230769230769231
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

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

```

An input tensor was not cuda.

```
warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
5742it [00:16, 355.06it/s]
load_f1 = 0.9275634995296332
real_f1 = 0.9274269557021678
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, 363.81it/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: 0.8228508830070496
```

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

```
step: 10, loss: 0.06579922884702682
```

```
step: 20, loss: 0.3784535527229309
```

```
step: 30, loss: 0.3717435896396637
```

```
step: 40, loss: 0.5045516490936279
```

```
step: 50, loss: 0.3000154495239258
```

```
step: 60, loss: 0.3527340590953827
```

```
step: 70, loss: 0.2319360226392746
```

```
step: 80, loss: 0.29881080985069275
```

```
step: 90, loss: 0.3951594829559326
```

```
step: 100, loss: 0.16337484121322632
```

```

step: 110, loss: 0.29786837100982666
step: 120, loss: 0.20926369726657867
step: 130, loss: 0.22394393384456635
step: 140, loss: 0.255533903837204
step: 150, loss: 0.2310546189546585
step: 160, loss: 0.14807860553264618
step: 170, loss: 0.1176474392414093
step: 180, loss: 0.17234131693840027
step: 190, loss: 0.2671186625957489
step: 200, loss: 0.13720162212848663
step: 210, loss: 0.3767639696598053
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5273069679849341, f1=0.5317460317460317,
best_f1=0.5317460317460317
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.056428294628858566
step: 10, loss: 0.10168114304542542
step: 20, loss: 0.21267978847026825
step: 30, loss: 0.07428642362356186
step: 40, loss: 0.11530536413192749
step: 50, loss: 0.2570505142211914
step: 60, loss: 0.04768776521086693
step: 70, loss: 0.16393797099590302
step: 80, loss: 0.1832418143749237
step: 90, loss: 0.1054469645023346
step: 100, loss: 0.06615502387285233
step: 110, loss: 0.07352042943239212
step: 120, loss: 0.14919568598270416
step: 130, loss: 0.16026903688907623
step: 140, loss: 0.25971749424934387
step: 150, loss: 0.18637143075466156
step: 160, loss: 0.1102469265460968
step: 170, loss: 0.11910095065832138
step: 180, loss: 0.29838302731513977
step: 190, loss: 0.23812387883663177
step: 200, loss: 0.1858142763376236
step: 210, loss: 0.19407224655151367
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.593186372745491, f1=0.626984126984127,
best_f1=0.626984126984127
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1514977514743805
step: 10, loss: 0.25698256492614746
step: 20, loss: 0.20889398455619812
step: 30, loss: 0.14225347340106964
step: 40, loss: 0.10756567120552063
step: 50, loss: 0.19958551228046417
step: 60, loss: 0.25472375750541687
step: 70, loss: 0.09323787689208984
step: 80, loss: 0.0886855274438858
step: 90, loss: 0.09174536913633347
step: 100, loss: 0.029307032003998756
step: 110, loss: 0.12508779764175415
step: 120, loss: 0.18233279883861542
step: 130, loss: 0.1694454699754715
step: 140, loss: 0.23054596781730652
step: 150, loss: 0.08986970037221909
step: 160, loss: 0.03799634799361229
step: 170, loss: 0.13957445323467255
step: 180, loss: 0.1507474035024643
step: 190, loss: 0.155556321144104
step: 200, loss: 0.08348597586154938
step: 210, loss: 0.15089625120162964
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5992779783393503, f1=0.6090090090090089,
best_f1=0.6090090090090089
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0762939304113388
step: 10, loss: 0.019046161323785782
step: 20, loss: 0.12050756812095642
step: 30, loss: 0.06204024329781532
step: 40, loss: 0.03241259977221489

```



```

step: 50, loss: 0.05478385463356972
step: 60, loss: 0.08544622361660004
step: 70, loss: 0.0912635549902916
step: 80, loss: 0.2041298896074295
step: 90, loss: 0.002137275179848075
step: 100, loss: 0.11455520987510681
step: 110, loss: 0.1277453452348709
step: 120, loss: 0.049431491643190384
step: 130, loss: 0.1978263556957245
step: 140, loss: 0.1582556664943695
step: 150, loss: 0.06711101531982422
step: 160, loss: 0.046644117683172226
step: 170, loss: 0.03273925557732582
step: 180, loss: 0.11055515706539154
step: 190, loss: 0.11365185678005219
step: 200, loss: 0.041853226721286774
step: 210, loss: 0.025252480059862137
/usr/local/lib/python3.7/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=0.5688405797101448, f1=0.5951557093425606,
best_f1=0.6090090090090089
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05016730725765228
step: 10, loss: 0.03467864170670509
step: 20, loss: 0.021461663767695427
step: 30, loss: 0.007512648589909077
step: 40, loss: 0.08440592139959335
step: 50, loss: 0.07277693599462509
step: 60, loss: 0.11556520313024521
step: 70, loss: 0.1298428177833557
step: 80, loss: 0.02145180106163025
step: 90, loss: 0.2054010033607483
step: 100, loss: 0.06803303956985474
step: 110, loss: 0.020925400778651237
step: 120, loss: 0.01142027135938406
step: 130, loss: 0.02260308526456356
step: 140, loss: 0.006820662412792444
step: 150, loss: 0.00956832803785801
step: 160, loss: 0.019934697076678276
step: 170, loss: 0.14677873253822327
step: 180, loss: 0.10475148260593414
step: 190, loss: 0.14043602347373962
step: 200, loss: 0.04328837990760803
step: 210, loss: 0.07413005828857422
/usr/local/lib/python3.7/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=0.5984848484848485, f1=0.6420664206642066,
best_f1=0.6090090090090089
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.033345527946949005
step: 10, loss: 0.018062923103570938
step: 20, loss: 0.020607581362128258
step: 30, loss: 0.027569513767957687
step: 40, loss: 0.0044368719682097435
step: 50, loss: 0.0943320095539093
step: 60, loss: 0.19020046293735504
step: 70, loss: 0.05906891077756882
step: 80, loss: 0.14519187808036804
step: 90, loss: 0.15001855790615082
step: 100, loss: 0.04351099953055382
step: 110, loss: 0.12797576189041138
step: 120, loss: 0.016328204423189163
step: 130, loss: 0.04405634477734566
step: 140, loss: 0.10170374065637589
step: 150, loss: 0.13718874752521515
step: 160, loss: 0.040188781917095184
step: 170, loss: 0.006189368665218353
step: 180, loss: 0.0021438151597976685
step: 190, loss: 0.029883284121751785
step: 200, loss: 0.026155229657888412
step: 210, loss: 0.004575734958052635
/usr/local/lib/python3.7/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=0.5934065934065933, f1=0.5896980461811723,
best_f1=0.6090090090090089
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.011238763108849525
step: 10, loss: 0.014391643926501274
step: 20, loss: 0.02840413711965084
step: 30, loss: 0.031197119504213333
step: 40, loss: 0.04293720796704292
step: 50, loss: 0.036034125834703445
step: 60, loss: 0.08636750280857086
step: 70, loss: 0.0024246848188340664
step: 80, loss: 0.012205083854496479
step: 90, loss: 0.03654209151864052
step: 100, loss: 0.024842550978064537

```

```

step: 110, loss: 0.006345321889966726
step: 120, loss: 0.27370694279670715
step: 130, loss: 0.0063856602646410465
step: 140, loss: 0.0031255281064659357
step: 150, loss: 0.012349222786724567
step: 160, loss: 0.007544314488768578
step: 170, loss: 0.0170072503387928
step: 180, loss: 0.0008064259891398251
step: 190, loss: 0.130354642868042
step: 200, loss: 0.0159634780138731
step: 210, loss: 0.041635140776634216
/usr/local/lib/python3.7/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=0.5720164609053499, f1=0.6061776061776061,
best_f1=0.6090090090090089
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02999930828809738
step: 10, loss: 0.03063712827861309
step: 20, loss: 0.04943997785449028
step: 30, loss: 0.0026711251121014357
step: 40, loss: 0.004715390503406525
step: 50, loss: 0.011497951112687588
step: 60, loss: 0.18922382593154907
step: 70, loss: 0.0020780526101589203
step: 80, loss: 0.002697278279811144
step: 90, loss: 0.02222507633268833
step: 100, loss: 0.001756322686560452
step: 110, loss: 0.010353689081966877
step: 120, loss: 0.015458413399755955
step: 130, loss: 0.005017285700887442
step: 140, loss: 0.003397604450583458
step: 150, loss: 0.00029935428756289184
step: 160, loss: 0.028104785829782486
step: 170, loss: 0.0013496587052941322
step: 180, loss: 0.06934347003698349
step: 190, loss: 0.048817168921232224
step: 200, loss: 0.07821021229028702
step: 210, loss: 0.03891732916235924
/usr/local/lib/python3.7/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=0.5794392523364487, f1=0.615658362989324,
best_f1=0.6090090090090089
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009923356585204601
step: 10, loss: 0.029627030715346336
step: 20, loss: 0.007118137553334236
step: 30, loss: 0.024601206183433533
step: 40, loss: 0.013039269484579563
step: 50, loss: 0.0017136135138571262
step: 60, loss: 0.0017005720874294639
step: 70, loss: 0.10817116498947144
step: 80, loss: 0.00659400736913085
step: 90, loss: 0.004055602476000786
step: 100, loss: 0.005628096405416727
step: 110, loss: 0.00959847029298544
step: 120, loss: 0.0022427483927458525
step: 130, loss: 0.0026225612964481115
step: 140, loss: 0.013715065084397793
step: 150, loss: 0.000844574358779937
step: 160, loss: 0.04178288206458092
step: 170, loss: 0.01697385311126709
step: 180, loss: 0.04034304991364479
step: 190, loss: 0.030816566199064255
step: 200, loss: 0.0421360619366169
step: 210, loss: 0.07475664466619492
/usr/local/lib/python3.7/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=0.5584415584415585, f1=0.5732217573221757,
best_f1=0.6090090090090089
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006822738330811262
step: 10, loss: 0.0022255759686231613
step: 20, loss: 0.01033520046621561
step: 30, loss: 0.0027493147645145655
step: 40, loss: 0.01581977680325508
step: 50, loss: 0.0003235328185837716
step: 60, loss: 0.0004248874611221254
step: 70, loss: 0.0019048380199819803
step: 80, loss: 0.00037854653783142567
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 90, loss: 0.0036286076065152884
step: 100, loss: 0.0004142956458963454
step: 110, loss: 0.000531049445271492
step: 120, loss: 0.008292239159345627
step: 130, loss: 0.0008272050181403756
step: 140, loss: 0.00025620960514061153
step: 150, loss: 0.027712181210517883

```

```

step: 160, loss: 0.06546398252248764
step: 170, loss: 0.0534069798886776
step: 180, loss: 0.0035833867732435465
step: 190, loss: 0.09409832954406738
step: 200, loss: 0.008327717892825603
step: 210, loss: 0.0005925953737460077
/usr/local/lib/python3.7/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=0.5756302521008404, f1=0.6196078431372549,
best_f1=0.6090090090090089
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0015951640671119094
step: 10, loss: 0.028734711930155754
step: 20, loss: 0.00033154129050672054
step: 30, loss: 0.055013615638017654
step: 40, loss: 0.0006285496638156474
step: 50, loss: 0.000515358115080744
step: 60, loss: 0.011388854123651981
step: 70, loss: 0.0011656066635623574
step: 80, loss: 0.01689479500055313
step: 90, loss: 0.0009967788355425
step: 100, loss: 0.010575132444500923
step: 110, loss: 0.08983069658279419
step: 120, loss: 0.0008898349478840828
step: 130, loss: 0.016038520261645317
step: 140, loss: 0.027401035651564598
step: 150, loss: 0.02890361286699772
step: 160, loss: 0.0017071746988222003
step: 170, loss: 0.18593072891235352
step: 180, loss: 0.023357214406132698
step: 190, loss: 0.0013978153001517057
step: 200, loss: 0.003346867859363556
step: 210, loss: 0.000604510132689029
/usr/local/lib/python3.7/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=0.5596707818930041, f1=0.6027397260273972,
best_f1=0.6090090090090089
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02002270333468914
step: 10, loss: 0.0008226231439039111
step: 20, loss: 0.024400237947702408
step: 30, loss: 0.00041776165016926825

```

```

step: 40, loss: 0.06716015189886093
step: 50, loss: 0.011610540561378002
step: 60, loss: 0.00033288620761595666
step: 70, loss: 0.0002770766441244632
step: 80, loss: 0.0015505511546507478
step: 90, loss: 0.00017360475612804294
step: 100, loss: 0.0023108553141355515
step: 110, loss: 0.0005915439105592668
step: 120, loss: 0.0006074330885894597
step: 130, loss: 0.004969288595020771
step: 140, loss: 0.0006952915573492646
step: 150, loss: 0.0002496030938345939
step: 160, loss: 0.012445573695003986
step: 170, loss: 0.008597364649176598
step: 180, loss: 0.0008096008677966893
step: 190, loss: 0.07064127177000046
step: 200, loss: 0.05366057902574539
step: 210, loss: 0.0008294355939142406
/usr/local/lib/python3.7/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=0.5542725173210161, f1=0.5855855855855856,
best_f1=0.6090090090090089
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001258292031707242
step: 10, loss: 0.002711046487092972
step: 20, loss: 0.0019861466716974974
step: 30, loss: 0.1368863731622696
step: 40, loss: 0.0002753433655016124
step: 50, loss: 0.00036294074379839003
step: 60, loss: 0.002476111985743046
step: 70, loss: 0.07150739431381226
step: 80, loss: 0.00046175686293281615
step: 90, loss: 0.17077048122882843
step: 100, loss: 0.00732789933681488
step: 110, loss: 0.00039827782893553376
step: 120, loss: 0.0160031970590353
step: 130, loss: 0.0021578159648925066
step: 140, loss: 0.0009047469357028604
step: 150, loss: 0.0010095451725646853
step: 160, loss: 0.0012724135303869843
step: 170, loss: 0.002124585211277008
step: 180, loss: 0.002567098941653967
step: 190, loss: 0.0002913328353315592
step: 200, loss: 0.00022836837160866708
step: 210, loss: 0.003250442212447524

```

```

/usr/local/lib/python3.7/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=0.5635593220338982, f1=0.6085192697768762,
best_f1=0.6090090090090089
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00389312906190753
step: 10, loss: 0.0003018052375409752
step: 20, loss: 0.00016712555952835828
step: 30, loss: 0.0004953111638315022
step: 40, loss: 0.0012169982073828578
step: 50, loss: 0.013007542118430138
step: 60, loss: 0.0056512863375246525
step: 70, loss: 0.005914785899221897
step: 80, loss: 0.006680551916360855
step: 90, loss: 0.0006485538906417787
step: 100, loss: 0.00042088067857548594
step: 110, loss: 0.007602238096296787
step: 120, loss: 0.005558631848543882
step: 130, loss: 0.0003243083774577826
step: 140, loss: 0.0011889170855283737
step: 150, loss: 0.006646659690886736
step: 160, loss: 0.0008530956692993641
step: 170, loss: 0.04169522970914841
step: 180, loss: 0.0006430522189475596
step: 190, loss: 0.00045624529593624175
step: 200, loss: 0.0002876415092032403
step: 210, loss: 0.027225099503993988
/usr/local/lib/python3.7/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=0.5478260869565217, f1=0.6058091286307054,
best_f1=0.6090090090090089
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002052883617579937
step: 10, loss: 0.0016537648625671864
step: 20, loss: 0.00015095339040271938
step: 30, loss: 0.0002305259695276618
step: 40, loss: 0.00020733311248477548
step: 50, loss: 0.00016256500384770334
step: 60, loss: 0.013453506864607334
step: 70, loss: 0.0004336564743425697
step: 80, loss: 0.0038509166333824396
step: 90, loss: 0.0003254881885368377

```

```

step: 100, loss: 0.00022995365725364536
step: 110, loss: 0.00010850894614122808
step: 120, loss: 0.00046081008622422814
step: 130, loss: 0.0020246696658432484
step: 140, loss: 0.0003213191812392324
step: 150, loss: 0.00017994080553762615
step: 160, loss: 0.00016150526062119752
step: 170, loss: 0.002108521293848753
step: 180, loss: 0.00234228209592402
step: 190, loss: 0.001034627901390195
step: 200, loss: 0.0002964668383356184
step: 210, loss: 0.0004879671905655414
/usr/local/lib/python3.7/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=0.5470852017937221, f1=0.6029723991507431,
best_f1=0.6090090090090089
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning:  multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
2293it [00:04, 560.34it/s]

```



```
load_f1 = 0.5946969696969697
real_f1 = 0.5877862595419848
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, 366.78it/s]
```

3.1.5 Walmart-Amazon - Running the matcher

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

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

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

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

Defaults for this optimization level are:
enabled : True

```

opt_level          : 02
cast_model_type    : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights     : True
loss_scale         : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled           : True
opt_level         : 02
cast_model_type   : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights    : True
loss_scale        : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8587543368339539
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.16086530685424805
step: 20, loss: 0.1552099883556366
step: 30, loss: 0.49749624729156494
step: 40, loss: 0.25086814165115356
step: 50, loss: 0.3125515580177307
step: 60, loss: 0.3652747869491577
step: 70, loss: 0.17115874588489532
step: 80, loss: 0.5166541934013367
step: 90, loss: 0.24345454573631287
step: 100, loss: 0.2192048728466034
step: 110, loss: 0.23718354105949402
step: 120, loss: 0.4142569899559021
step: 130, loss: 0.346358984708786
step: 140, loss: 0.29708781838417053
step: 150, loss: 0.2557724118232727

```

```

step: 160, loss: 0.2061873972415924
step: 170, loss: 0.30934005975723267
step: 180, loss: 0.30738934874534607
step: 190, loss: 0.09206560999155045
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6015831134564644, f1=0.6294416243654822,
best_f1=0.6294416243654822
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.24821701645851135
step: 10, loss: 0.027940204367041588
step: 20, loss: 0.08490920066833496
step: 30, loss: 0.13289372622966766
step: 40, loss: 0.48247525095939636
step: 50, loss: 0.31255170702934265
step: 60, loss: 0.2292994260787964
step: 70, loss: 0.16225525736808777
step: 80, loss: 0.18230122327804565
step: 90, loss: 0.11200247704982758
step: 100, loss: 0.2211366891860962
step: 110, loss: 0.15505412220954895
step: 120, loss: 0.1760723888874054
step: 130, loss: 0.08824549615383148
step: 140, loss: 0.05614502355456352
step: 150, loss: 0.023750649765133858
step: 160, loss: 0.08439387381076813
step: 170, loss: 0.13960067927837372
step: 180, loss: 0.20393110811710358
step: 190, loss: 0.1612468957901001
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7520891364902507,
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: 0.15666937828063965
step: 10, loss: 0.23656712472438812
step: 20, loss: 0.02991829812526703
step: 30, loss: 0.09321460872888565
step: 40, loss: 0.048635002225637436
step: 50, loss: 0.2878531813621521
step: 60, loss: 0.0575251579284668
step: 70, loss: 0.21419207751750946
step: 80, loss: 0.1660798192024231
step: 90, loss: 0.12365496903657913
step: 100, loss: 0.1262088268995285
step: 110, loss: 0.17111411690711975
step: 120, loss: 0.028435910120606422
step: 130, loss: 0.06481233984231949
step: 140, loss: 0.11394038796424866
step: 150, loss: 0.07844867557287216
step: 160, loss: 0.059744276106357574
step: 170, loss: 0.05911784991621971
step: 180, loss: 0.03445785865187645
step: 190, loss: 0.1684291660785675
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7236180904522613, f1=0.7696202531645571,
best_f1=0.7696202531645571
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.049832943826913834
step: 10, loss: 0.07271531224250793
step: 20, loss: 0.020521780475974083
step: 30, loss: 0.06701849400997162
step: 40, loss: 0.002946312539279461
step: 50, loss: 0.011350937187671661
step: 60, loss: 0.030442550778388977
step: 70, loss: 0.0033010817132890224
step: 80, loss: 0.005555982701480389
step: 90, loss: 0.01627620868384838
step: 100, loss: 0.010107059963047504
step: 110, loss: 0.01849178597331047
step: 120, loss: 0.09309488534927368
step: 130, loss: 0.20414143800735474
step: 140, loss: 0.08325020968914032
step: 150, loss: 0.0012993001146242023

```

```

step: 160, loss: 0.018959177657961845
step: 170, loss: 0.05093414708971977
step: 180, loss: 0.11461175233125687
step: 190, loss: 0.01962977647781372
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7383863080684596, f1=0.7475728155339805,
best_f1=0.7475728155339805
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.019200321286916733
step: 10, loss: 0.020388619974255562
step: 20, loss: 0.1239968091249466
step: 30, loss: 0.004531236831098795
step: 40, loss: 0.020365087315440178
step: 50, loss: 0.012536230497062206
step: 60, loss: 0.022953476756811142
step: 70, loss: 0.0007097611087374389
step: 80, loss: 0.0033099562861025333
step: 90, loss: 0.011245504952967167
step: 100, loss: 0.1244858056306839
step: 110, loss: 0.006138361059129238
step: 120, loss: 0.0013769448269158602
step: 130, loss: 0.01160522736608982
step: 140, loss: 0.011780896224081516
step: 150, loss: 0.01177866943180561
step: 160, loss: 0.0050571877509355545
step: 170, loss: 0.019163809716701508
step: 180, loss: 0.09644801914691925
step: 190, loss: 0.023135975003242493
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7553191489361701, f1=0.746031746031746,
best_f1=0.746031746031746
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.07454963028430939
step: 10, loss: 0.0015097062569111586
step: 20, loss: 0.035785481333732605
step: 30, loss: 0.0017437342321500182
step: 40, loss: 0.005899276118725538
step: 50, loss: 0.001646126271225512
step: 60, loss: 0.03023318573832512
step: 70, loss: 0.01959429495036602
step: 80, loss: 0.10142218321561813
step: 90, loss: 0.07235976308584213
step: 100, loss: 0.026000220328569412
step: 110, loss: 0.04165026172995567
step: 120, loss: 0.0034042035695165396
step: 130, loss: 0.002338594989851117
step: 140, loss: 0.008499088697135448
step: 150, loss: 0.013363123871386051
step: 160, loss: 0.001595636480487883
step: 170, loss: 0.11505208909511566
step: 180, loss: 0.0428125225007534
step: 190, loss: 0.01929900422692299
/usr/local/lib/python3.7/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=0.7234042553191489, f1=0.7561643835616437,
best_f1=0.746031746031746
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0014806005638092756
step: 10, loss: 0.014886241406202316
step: 20, loss: 0.0005062426789663732
step: 30, loss: 0.025636469945311546
step: 40, loss: 0.0034387102350592613
step: 50, loss: 0.07427891343832016
step: 60, loss: 0.0012827971950173378
step: 70, loss: 0.0038096955977380276
step: 80, loss: 0.043953992426395416
step: 90, loss: 0.008828014135360718
step: 100, loss: 0.0034083030186593533
step: 110, loss: 0.012170298025012016
step: 120, loss: 0.0015393313951790333
step: 130, loss: 0.000522148737218231
step: 140, loss: 0.01660427823662758
step: 150, loss: 0.0017905068816617131
step: 160, loss: 0.010425865650177002
step: 170, loss: 0.032778237015008926
step: 180, loss: 0.002399219200015068
step: 190, loss: 0.017211178317666054

```

```

/usr/local/lib/python3.7/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=0.7253333333333334, f1=0.7258064516129032,
best_f1=0.746031746031746
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009515408310107887
step: 10, loss: 0.01557355560362339
step: 20, loss: 0.0011472213082015514
step: 30, loss: 0.00299431593157351
step: 40, loss: 0.004282744135707617
step: 50, loss: 0.0015382957644760609
step: 60, loss: 0.0031179965008050203
step: 70, loss: 0.0035353933926671743
step: 80, loss: 0.004232273902744055
step: 90, loss: 0.010377822443842888
step: 100, loss: 0.01668674312531948
step: 110, loss: 0.016128258779644966
step: 120, loss: 0.04917436093091965
step: 130, loss: 0.005052416585385799
step: 140, loss: 0.011699255555868149
step: 150, loss: 0.0010275394888594747
step: 160, loss: 0.002233152510598302
step: 170, loss: 0.0004225385491736233
step: 180, loss: 0.002718034666031599
step: 190, loss: 0.00263771740719676
/usr/local/lib/python3.7/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=0.7282913165266107, f1=0.7485714285714286,
best_f1=0.746031746031746
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002637428930029273
step: 10, loss: 0.02139655500650406
step: 20, loss: 0.036144357174634933
step: 30, loss: 0.046251844614744186
step: 40, loss: 0.0021080656442791224
step: 50, loss: 0.0009967482183128595
step: 60, loss: 0.0013201807159930468
step: 70, loss: 0.010717116296291351
step: 80, loss: 0.0025601848028600216
step: 90, loss: 0.0021401294507086277
step: 100, loss: 0.0028644385747611523
step: 110, loss: 0.0007418652530759573

```

```

step: 120, loss: 0.0017630669753998518
step: 130, loss: 0.011427109129726887
step: 140, loss: 0.0003210093709640205
step: 150, loss: 0.00170648202765733
step: 160, loss: 0.0006637285696342587
step: 170, loss: 0.008562169037759304
step: 180, loss: 0.144056037068367
step: 190, loss: 0.0017736316658556461
/usr/local/lib/python3.7/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=0.7202216066481995, f1=0.721763085399449,
best_f1=0.746031746031746
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002882011467590928
step: 10, loss: 0.0024105445481836796
step: 20, loss: 0.0006620128988288343
step: 30, loss: 0.06440150737762451
step: 40, loss: 0.00031028370722196996
step: 50, loss: 0.0003417872649151832
step: 60, loss: 0.001752282027155161
step: 70, loss: 0.0014760267222300172
step: 80, loss: 0.00023285544011741877
step: 90, loss: 0.000970548193436116
step: 100, loss: 0.0039001430850476027
step: 110, loss: 0.0019694173242896795
step: 120, loss: 0.0020508153829723597
step: 130, loss: 0.0007603639387525618
step: 140, loss: 0.0004001236811745912
step: 150, loss: 0.00032652844674885273
step: 160, loss: 0.00041215596138499677
step: 170, loss: 0.0014367711264640093
step: 180, loss: 0.0010669903131201863
step: 190, loss: 0.0006393804214894772
/usr/local/lib/python3.7/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=0.7371007371007371, f1=0.7146401985111663,
best_f1=0.746031746031746
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003577528987079859
step: 10, loss: 0.003214135766029358
step: 20, loss: 0.0011522144777700305
step: 30, loss: 0.001685017254203558

```



```

step: 40, loss: 0.004364004824310541
step: 50, loss: 0.001597968628630042
step: 60, loss: 0.00021732020832132548
step: 70, loss: 0.0015023209853097796
step: 80, loss: 0.0001480002247262746
step: 90, loss: 0.00018772517796605825
step: 100, loss: 0.00029538641683757305
step: 110, loss: 0.00026859581703320146
step: 120, loss: 0.0002749319828581065
step: 130, loss: 0.00037362350849434733
step: 140, loss: 0.00012185872037662193
step: 150, loss: 0.002282602246850729
step: 160, loss: 0.0001690241479082033
step: 170, loss: 0.00020168998162262142
step: 180, loss: 0.001273180590942502
step: 190, loss: 0.0007293369853869081
/usr/local/lib/python3.7/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=0.7355163727959698, f1=0.7379134860050889,
best_f1=0.746031746031746
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00030516329570673406
step: 10, loss: 0.004620726685971022
step: 20, loss: 0.0013657953822985291
step: 30, loss: 0.0034568295814096928
step: 40, loss: 0.0003471620730124414
step: 50, loss: 0.0010426489170640707
step: 60, loss: 0.0017679346492514014
step: 70, loss: 0.00022959389025345445
step: 80, loss: 0.0011685595382004976
step: 90, loss: 0.0020306778606027365
step: 100, loss: 0.00041165651055052876
step: 110, loss: 9.693895117379725e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 120, loss: 0.00045051914639770985
step: 130, loss: 0.00016816638526506722
step: 140, loss: 0.0038601511623710394
step: 150, loss: 0.0006781858974136412
step: 160, loss: 0.000890817609615624
step: 170, loss: 0.0007703368319198489
step: 180, loss: 0.00031550059793516994
step: 190, loss: 0.00029990196344442666
/usr/local/lib/python3.7/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=0.7418546365914787, f1=0.745, best_f1=0.746031746031746
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00015069165965542197
step: 10, loss: 0.0002617222198750824
step: 20, loss: 0.0009499791776761413
step: 30, loss: 0.0001956918422365561
step: 40, loss: 0.00012062558380421251
step: 50, loss: 0.0012244904646649957
step: 60, loss: 0.00018909721984528005
step: 70, loss: 0.0009241189109161496
step: 80, loss: 0.0005491807241924107
step: 90, loss: 0.0006099400925450027
step: 100, loss: 0.001638164627365768
step: 110, loss: 0.0007396938162855804
step: 120, loss: 0.005213105119764805
step: 130, loss: 0.015852564945816994
step: 140, loss: 0.0005016109789721668
step: 150, loss: 9.37549993977882e-05
step: 160, loss: 0.00023329138639383018
step: 170, loss: 0.00022468474344350398
step: 180, loss: 0.0010286688338965178
step: 190, loss: 0.0010637366212904453
/usr/local/lib/python3.7/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=0.7403598971722366, f1=0.744186046511628,
best_f1=0.746031746031746
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009173059370368719
step: 10, loss: 0.0006815666565671563
step: 20, loss: 0.0013327200431376696
step: 30, loss: 0.000785522279329598
step: 40, loss: 0.00030880304984748363
step: 50, loss: 0.08385465294122696
step: 60, loss: 0.0028610110748559237
step: 70, loss: 0.00039674874278716743
step: 80, loss: 8.700120815774426e-05
step: 90, loss: 0.0020387943368405104
step: 100, loss: 0.0003267930878791958
step: 110, loss: 0.0003452039381954819
step: 120, loss: 0.0004019327461719513
step: 130, loss: 0.005925491917878389
step: 140, loss: 0.0001495396572863683
step: 150, loss: 0.0004947272245772183

```

```

step: 160, loss: 0.0004708691849373281
step: 170, loss: 0.0002470817125868052
step: 180, loss: 0.0007375019486062229
step: 190, loss: 0.00019368136418052018
/usr/local/lib/python3.7/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=0.7351351351351353, f1=0.7433155080213903,
best_f1=0.746031746031746
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00023987512395251542
step: 10, loss: 0.00014274120621848851
step: 20, loss: 0.0004703309678006917
step: 30, loss: 0.001979873050004244
step: 40, loss: 0.009323443286120892
step: 50, loss: 0.00013879583275411278
step: 60, loss: 0.0009815822122618556
step: 70, loss: 0.0003367029130458832
step: 80, loss: 0.0013847114751115441
step: 90, loss: 0.000905253691598773
step: 100, loss: 0.0001627097954042256
step: 110, loss: 0.0017394623719155788
step: 120, loss: 0.00026898563373833895
step: 130, loss: 0.0004221608687657863
step: 140, loss: 0.0029986072331666946
step: 150, loss: 0.00022960180649533868
step: 160, loss: 0.00034449753002263606
step: 170, loss: 0.0002475257497280836
step: 180, loss: 0.003058216068893671
step: 190, loss: 0.004297089297324419
/usr/local/lib/python3.7/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=0.7331536388140162, f1=0.7446808510638299,
best_f1=0.746031746031746
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

```

Defaults for this optimization level are:

```

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

```

```

loss_scale          : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled             : True
opt_level           : 02
cast_model_type     : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32  : True
master_weights      : True
loss_scale          : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2049it [00:06, 321.74it/s]
load_f1 = 0.627027027027027
real_f1 = 0.6243386243386243
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, 373.32it/s]

```

3.2 DK TEXTUAL

3.2.1 Abt-Buy - Running the matcher

```

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

```

```

--dk product \
--summarize

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

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

```

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

Defaults for this optimization level are:

```

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

```

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

After processing overrides, optimization options are:

```

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

```

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

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

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

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

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

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

step: 0, loss: 0.8535327911376953

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

step: 10, loss: 0.22527579963207245

step: 20, loss: 0.1572115123271942

step: 30, loss: 0.2295931875705719

step: 40, loss: 0.31416893005371094

step: 50, loss: 0.3744119703769684

step: 60, loss: 0.4330957233905792

step: 70, loss: 0.30707499384880066

step: 80, loss: 0.24992646276950836

step: 90, loss: 0.403587281703949

step: 100, loss: 0.21970438957214355

step: 110, loss: 0.17741452157497406

step: 120, loss: 0.574905514717102

step: 130, loss: 0.40258434414863586

step: 140, loss: 0.461958110332489

step: 150, loss: 0.12606681883335114

step: 160, loss: 0.3080752491950989

step: 170, loss: 0.13382838666439056

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

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

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

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=0.625, f1=0.642512077294686, best_f1=0.642512077294686

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

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

step: 0, loss: 0.2602345645427704

step: 10, loss: 0.08939461410045624

step: 20, loss: 0.4422665238380432

step: 30, loss: 0.24342089891433716

step: 40, loss: 0.12124888598918915

step: 50, loss: 0.2564605474472046

step: 60, loss: 0.06202420964837074

step: 70, loss: 0.21101577579975128

step: 80, loss: 0.14158359169960022

step: 90, loss: 0.27525049448013306

step: 100, loss: 0.12768927216529846

step: 110, loss: 0.21013721823692322

```

step: 120, loss: 0.05146493390202522
step: 130, loss: 0.0809638500213623
step: 140, loss: 0.12882979214191437
step: 150, loss: 0.10011211782693863
step: 160, loss: 0.0762130469083786
step: 170, loss: 0.03802881017327309
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7586206896551724, f1=0.753880266075388,
best_f1=0.753880266075388
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.058955248445272446
step: 10, loss: 0.07270509004592896
step: 20, loss: 0.0647953674197197
step: 30, loss: 0.1606401950120926
step: 40, loss: 0.008091687224805355
step: 50, loss: 0.16773107647895813
step: 60, loss: 0.07308939844369888
step: 70, loss: 0.07751382142305374
step: 80, loss: 0.18494316935539246
step: 90, loss: 0.07747131586074829
step: 100, loss: 0.04634484648704529
step: 110, loss: 0.08836309611797333
step: 120, loss: 0.02572619356215
step: 130, loss: 0.12851224839687347
step: 140, loss: 0.013438225723803043
step: 150, loss: 0.2413737177848816
step: 160, loss: 0.04137277230620384
step: 170, loss: 0.12638190388679504
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7718446601941746, f1=0.7685185185185185,
best_f1=0.7685185185185185
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.1396910697221756
step: 10, loss: 0.17600777745246887
step: 20, loss: 0.04859861731529236
step: 30, loss: 0.015016536228358746
step: 40, loss: 0.1373586654663086
step: 50, loss: 0.017411038279533386
step: 60, loss: 0.03672143444418907
step: 70, loss: 0.0708707869052887
step: 80, loss: 0.17913897335529327
step: 90, loss: 0.02025524340569973
step: 100, loss: 0.009490590542554855
step: 110, loss: 0.007656867150217295
step: 120, loss: 0.16776055097579956
step: 130, loss: 0.06151023507118225
step: 140, loss: 0.007794760633260012
step: 150, loss: 0.037099581211805344
step: 160, loss: 0.027203144505620003
step: 170, loss: 0.10493266582489014
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.779746835443038, f1=0.7439613526570048,
best_f1=0.7439613526570048
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07860086113214493
step: 10, loss: 0.0185343399643898
step: 20, loss: 0.009103014133870602
step: 30, loss: 0.03669581189751625
step: 40, loss: 0.005042325239628553
step: 50, loss: 0.005357610527426004
step: 60, loss: 0.0007540574297308922
step: 70, loss: 0.0033629974350333214
step: 80, loss: 0.00938765611499548
step: 90, loss: 0.02147156558930874
step: 100, loss: 0.060843221843242645
step: 110, loss: 0.1339816153049469
step: 120, loss: 0.023104730993509293
step: 130, loss: 0.014610578306019306
step: 140, loss: 0.18152061104774475
step: 150, loss: 0.00807980541139841
step: 160, loss: 0.015437784604728222
step: 170, loss: 0.06067150831222534

```



```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7934272300469484, f1=0.7671232876712328,
best_f1=0.7671232876712328
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01889745518565178
step: 10, loss: 0.06346837431192398
step: 20, loss: 0.009129021316766739
step: 30, loss: 0.020741544663906097
step: 40, loss: 0.002169463550671935
step: 50, loss: 0.011589586734771729
step: 60, loss: 0.03394502028822899
step: 70, loss: 0.003928090445697308
step: 80, loss: 0.08314169198274612
step: 90, loss: 0.021931970492005348
step: 100, loss: 0.018914997577667236
step: 110, loss: 0.029352908954024315
step: 120, loss: 0.007623627781867981
step: 130, loss: 0.1973653882741928
step: 140, loss: 0.020735645666718483
step: 150, loss: 0.18360672891139984
step: 160, loss: 0.019923172891139984
step: 170, loss: 0.004510032013058662
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8144578313253013, f1=0.7713625866050808,
best_f1=0.7713625866050808
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008525331504642963
step: 10, loss: 0.0023758085444569588
step: 20, loss: 0.030850818380713463
step: 30, loss: 0.0034889006055891514
step: 40, loss: 0.007236321456730366
step: 50, loss: 0.002637225203216076

```

```

step: 60, loss: 0.07829312235116959
step: 70, loss: 0.025261567905545235
step: 80, loss: 0.0014173266245052218
step: 90, loss: 0.021122856065630913
step: 100, loss: 0.02221830002963543
step: 110, loss: 0.00015238850028254092
step: 120, loss: 0.11723446100950241
step: 130, loss: 0.1773279458284378
step: 140, loss: 0.004383209627121687
step: 150, loss: 0.017466774210333824
step: 160, loss: 0.006476543378084898
step: 170, loss: 0.12741348147392273
/usr/local/lib/python3.7/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=0.7989821882951653, f1=0.7850467289719626,
best_f1=0.7713625866050808
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.045072704553604126
step: 10, loss: 0.008523957803845406
step: 20, loss: 0.003465939313173294
step: 30, loss: 0.031633514910936356
step: 40, loss: 0.03311879187822342
step: 50, loss: 0.025452522560954094
step: 60, loss: 0.00488915154710412
step: 70, loss: 0.020877577364444733
step: 80, loss: 0.0037336174864321947
step: 90, loss: 0.010573441162705421
step: 100, loss: 0.0020005700644105673
step: 110, loss: 0.0055761332623660564
step: 120, loss: 0.052235376089811325
step: 130, loss: 0.01744077540934086
step: 140, loss: 0.0008435212075710297
step: 150, loss: 0.06697254627943039
step: 160, loss: 0.03660018742084503
step: 170, loss: 0.006106024142354727
/usr/local/lib/python3.7/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=0.8102564102564102, f1=0.7853658536585365,
best_f1=0.7713625866050808
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.010616451501846313
step: 10, loss: 0.005164664704352617

```

```

step: 20, loss: 0.005453967489302158
step: 30, loss: 0.038843680173158646
step: 40, loss: 0.002954735653474927
step: 50, loss: 0.014459570869803429
step: 60, loss: 0.00533245038241148
step: 70, loss: 0.00501618068665266
step: 80, loss: 0.1105261892080307
step: 90, loss: 0.005607341881841421
step: 100, loss: 0.028126852586865425
step: 110, loss: 9.132825653068721e-05
step: 120, loss: 0.0011704039061442018
step: 130, loss: 0.003274464514106512
step: 140, loss: 0.0009614668088033795
step: 150, loss: 0.012750640511512756
step: 160, loss: 0.09207269549369812
step: 170, loss: 0.020714085549116135
/usr/local/lib/python3.7/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=0.8080357142857143, f1=0.7593360995850623,
best_f1=0.7713625866050808
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0017129022162407637
step: 10, loss: 0.006888733711093664
step: 20, loss: 0.003440332366153598
step: 30, loss: 0.0004074389871675521
step: 40, loss: 0.0013682732824236155
step: 50, loss: 0.005444458685815334
step: 60, loss: 0.002280113287270069
step: 70, loss: 0.002756224013864994
step: 80, loss: 0.006157928612083197
step: 90, loss: 0.07647860795259476
step: 100, loss: 0.002278817817568779
step: 110, loss: 0.004661100450903177
step: 120, loss: 0.004486100282520056
step: 130, loss: 0.005556856747716665
step: 140, loss: 0.01054797787219286
step: 150, loss: 0.0007987456046976149
step: 160, loss: 0.0003865005564875901
step: 170, loss: 0.0009344096761196852
/usr/local/lib/python3.7/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=0.8069306930693069, f1=0.7757009345794393,
best_f1=0.7713625866050808
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008958036778494716
step: 10, loss: 0.004509772174060345
step: 20, loss: 0.08125561475753784
step: 30, loss: 0.0023300047032535076
step: 40, loss: 0.0026975974906235933
step: 50, loss: 0.011961468495428562
step: 60, loss: 0.06983120739459991
step: 70, loss: 0.022388599812984467
step: 80, loss: 0.0003980644978582859
step: 90, loss: 0.0019605110865086317
step: 100, loss: 0.000587400805670768
step: 110, loss: 0.01101861335337162
step: 120, loss: 0.0031674783676862717
step: 130, loss: 0.009576905518770218
step: 140, loss: 0.049317218363285065
step: 150, loss: 0.002781857270747423
step: 160, loss: 0.000838443695101887
step: 170, loss: 0.06940687447786331
/usr/local/lib/python3.7/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=0.7918781725888325, f1=0.7751196172248804,
best_f1=0.7713625866050808
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04115238040685654
step: 10, loss: 0.01666354574263096
step: 20, loss: 0.010973029769957066
step: 30, loss: 0.0012494155671447515
step: 40, loss: 0.00019213413179386407
step: 50, loss: 0.002249498385936022
step: 60, loss: 0.0008268608944490552
step: 70, loss: 0.057297516614198685
step: 80, loss: 0.012278771959245205
step: 90, loss: 0.00019741100550163537
step: 100, loss: 0.00309701357036829
step: 110, loss: 0.0003426472540013492
step: 120, loss: 0.00033344855182804167
step: 130, loss: 0.0005055404035374522
step: 140, loss: 0.0010430720867589116
step: 150, loss: 0.0006139901233837008
step: 160, loss: 0.00012755785428453237
step: 170, loss: 0.26701244711875916
/usr/local/lib/python3.7/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=0.8020565552699228, f1=0.7651331719128328,
best_f1=0.7713625866050808
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001844810030888766
step: 10, loss: 0.08297406882047653
step: 20, loss: 0.00026542021078057587
step: 30, loss: 0.00020764314103871584
step: 40, loss: 0.00016954896273091435
step: 50, loss: 0.0010974777396768332
step: 60, loss: 0.0005288460524752736
step: 70, loss: 0.005875056143850088
step: 80, loss: 0.0004896890022791922
step: 90, loss: 0.0001779076410457492
step: 100, loss: 0.00016128047718666494
step: 110, loss: 0.00021580478642135859
step: 120, loss: 0.00015788796008564532
step: 130, loss: 0.002524027368053794
step: 140, loss: 0.00863446295261383
step: 150, loss: 0.0003045399789698422
step: 160, loss: 0.0004041825304739177
step: 170, loss: 0.0011544175213202834
/usr/local/lib/python3.7/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=0.8050632911392405, f1=0.7751196172248804,
best_f1=0.7713625866050808
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003227945417165756
step: 10, loss: 0.0009943200275301933
step: 20, loss: 0.012775269336998463
step: 30, loss: 0.0006869298522360623
step: 40, loss: 0.00039747479604557157
step: 50, loss: 0.0007263199659064412
step: 60, loss: 0.00028981658397242427
step: 70, loss: 0.011873342096805573
step: 80, loss: 0.0005424490664154291
step: 90, loss: 0.005751639138907194
step: 100, loss: 0.0030767719727009535
step: 110, loss: 0.00266246753744781
step: 120, loss: 0.000587679271120578
step: 130, loss: 0.0001915246102726087
step: 140, loss: 0.004336018115282059
step: 150, loss: 0.002937508746981621

```

```

step: 160, loss: 0.039940427988767624
step: 170, loss: 0.00022232389892451465
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
epoch 14: dev_f1=0.8129675810473815, f1=0.7627906976744185,
best_f1=0.7713625866050808
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000613206357229501
step: 10, loss: 0.001377012813463807
step: 20, loss: 0.010004709474742413
step: 30, loss: 0.006253522355109453
step: 40, loss: 0.0003679132496472448
step: 50, loss: 0.0018421708373352885
step: 60, loss: 0.00013769477664027363
step: 70, loss: 0.015923621132969856
step: 80, loss: 0.00014038773952051997
step: 90, loss: 0.00012875373067799956
step: 100, loss: 0.0013560279039666057
step: 110, loss: 0.00043570532579906285
step: 120, loss: 0.02027173340320587
step: 130, loss: 0.033103179186582565
step: 140, loss: 0.00015149280079640448
step: 150, loss: 0.05496003478765488
step: 160, loss: 0.00018195691518485546
step: 170, loss: 0.000194371648831293
/usr/local/lib/python3.7/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=0.8120300751879699, f1=0.7688679245283018,
best_f1=0.7713625866050808
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

```

Defaults for this optimization level are:

```

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

```

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

After processing overrides, optimization options are:

```

enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
1916it [00:05, 368.92it/s]
load_f1 = 0.6157024793388429
real_f1 = 0.6134969325153374
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, 362.73it/s]

```

3.3 DK DIRTY

3.3.1 DBLP-ACM - Running the matcher

```

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

```

```

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

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

```

Selected optimization level 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: 0.7987403869628906
step: 10, loss: 0.4626925587654114
step: 20, loss: 0.5442101359367371
step: 30, loss: 0.36535248160362244
step: 40, loss: 0.17335261404514313
step: 50, loss: 0.11264615505933762
step: 60, loss: 0.16595333814620972
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 70, loss: 0.1373388022184372
step: 80, loss: 0.22668743133544922
step: 90, loss: 0.05684578791260719
step: 100, loss: 0.1902245283126831
step: 110, loss: 0.10345470160245895
step: 120, loss: 0.02688564546406269
step: 130, loss: 0.012215152382850647
step: 140, loss: 0.018430408090353012
step: 150, loss: 0.0977155938744545
step: 160, loss: 0.26050227880477905
step: 170, loss: 0.03108414076268673
step: 180, loss: 0.027705999091267586
step: 190, loss: 0.03820079192519188
step: 200, loss: 0.005596744827926159
step: 210, loss: 0.007641276344656944
step: 220, loss: 0.010200342163443565
step: 230, loss: 0.03164626285433769
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9587513935340021, f1=0.954954954954955,
best_f1=0.954954954954955
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.056247394531965256
step: 10, loss: 0.002624292392283678
step: 20, loss: 0.004611840471625328
step: 30, loss: 0.00483226403594017
step: 40, loss: 0.04991960525512695
step: 50, loss: 0.04209453612565994
step: 60, loss: 0.01329534500837326
step: 70, loss: 0.009270762093365192

```

```

step: 80, loss: 0.004613013472408056
step: 90, loss: 0.0324602834880352
step: 100, loss: 0.1244128867983818
step: 110, loss: 0.15572798252105713
step: 120, loss: 0.007952834479510784
step: 130, loss: 0.04558971896767616
step: 140, loss: 0.11673150956630707
step: 150, loss: 0.015538297593593597
step: 160, loss: 0.013865937478840351
step: 170, loss: 0.04557831585407257
step: 180, loss: 0.015953153371810913
step: 190, loss: 0.07297901809215546
step: 200, loss: 0.0178364310413599
step: 210, loss: 0.1354529708623886
step: 220, loss: 0.009079154580831528
step: 230, loss: 0.023124555125832558
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9707865168539327, f1=0.9729119638826186,
best_f1=0.9729119638826186
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04682132229208946
step: 10, loss: 0.007033246103674173
step: 20, loss: 0.0066938744857907295
step: 30, loss: 0.031392090022563934
step: 40, loss: 0.10757895559072495
step: 50, loss: 0.004344021901488304
step: 60, loss: 0.0257562268525362
step: 70, loss: 0.005012055858969688
step: 80, loss: 0.03129057586193085
step: 90, loss: 0.051822103559970856
step: 100, loss: 0.004282364156097174
step: 110, loss: 0.011432324536144733
step: 120, loss: 0.017282243818044662
step: 130, loss: 0.007069905288517475
step: 140, loss: 0.0466017983853817
step: 150, loss: 0.007004712708294392
step: 160, loss: 0.0026028372813016176
step: 170, loss: 0.0025113008450716734
step: 180, loss: 0.005200927145779133
step: 190, loss: 0.008833430707454681

```

```

step: 200, loss: 0.018491078168153763
step: 210, loss: 0.012494838796555996
step: 220, loss: 0.01256814319640398
step: 230, loss: 0.07164563238620758
/usr/local/lib/python3.7/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=0.9692832764505119, f1=0.9658314350797267,
best_f1=0.9729119638826186
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003942964132875204
step: 10, loss: 0.00303676538169384
step: 20, loss: 0.0008109656046144664
step: 30, loss: 0.0016334418905898929
step: 40, loss: 0.0045339749194681644
step: 50, loss: 0.0020242577884346247
step: 60, loss: 0.0019870088435709476
step: 70, loss: 0.06281188875436783
step: 80, loss: 0.16132624447345734
step: 90, loss: 0.002864836249500513
step: 100, loss: 0.0022662749979645014
step: 110, loss: 0.01361505314707756
step: 120, loss: 0.0033921804279088974
step: 130, loss: 0.04942706599831581
step: 140, loss: 0.00875425711274147
step: 150, loss: 0.011640856973826885
step: 160, loss: 0.005198382306843996
step: 170, loss: 0.007156578358262777
step: 180, loss: 0.094502292573452
step: 190, loss: 0.010714694857597351
step: 200, loss: 0.000987350125797093
step: 210, loss: 0.002015827689319849
step: 220, loss: 0.0015077150892466307
step: 230, loss: 0.0006199266063049436
/usr/local/lib/python3.7/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=0.9679558011049725, f1=0.967525195968645,
best_f1=0.9729119638826186
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0007578032091259956
step: 10, loss: 0.0019295306410640478
step: 20, loss: 0.0004899029154330492
step: 30, loss: 0.000474881031550467

```

```

step: 40, loss: 0.00037865768535993993
step: 50, loss: 0.002061479724943638
step: 60, loss: 0.001014119479805231
step: 70, loss: 0.0075168912298977375
step: 80, loss: 0.0011957800015807152
step: 90, loss: 0.001415829174220562
step: 100, loss: 0.004180117975920439
step: 110, loss: 0.001555651775561273
step: 120, loss: 0.035364482551813126
step: 130, loss: 0.018636394292116165
step: 140, loss: 0.01350193191319704
step: 150, loss: 0.001254764967598021
step: 160, loss: 0.1405196189880371
step: 170, loss: 0.0318402424454689
step: 180, loss: 0.0013141692616045475
step: 190, loss: 0.0008219928713515401
step: 200, loss: 0.003967662807554007
step: 210, loss: 0.0013059901539236307
step: 220, loss: 0.010036949068307877
step: 230, loss: 0.0072493962943553925
/usr/local/lib/python3.7/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=0.9696287964004499, f1=0.9642058165548099,
best_f1=0.9729119638826186
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014078213833272457
step: 10, loss: 0.014984291046857834
step: 20, loss: 0.004464081954210997
step: 30, loss: 0.0008137645781971514
step: 40, loss: 0.0017009059665724635
step: 50, loss: 0.029145702719688416
step: 60, loss: 0.007521835155785084
step: 70, loss: 0.0035520452074706554
step: 80, loss: 0.008240998722612858
step: 90, loss: 0.0013926841784268618
step: 100, loss: 0.000524464063346386
step: 110, loss: 0.044831305742263794
step: 120, loss: 0.005006460938602686
step: 130, loss: 0.003932002000510693
step: 140, loss: 0.0008309084223583341
step: 150, loss: 0.00035002894583158195
step: 160, loss: 0.1281866729259491
step: 170, loss: 0.0018588254461064935
step: 180, loss: 0.0018174824072048068
step: 190, loss: 0.012898187153041363

```

```

step: 200, loss: 0.009245531633496284
step: 210, loss: 0.0011153888190165162
step: 220, loss: 0.0008590822690166533
step: 230, loss: 0.0044526392593979836
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9776286353467561, f1=0.9744160177975528,
best_f1=0.9744160177975528
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03466716408729553
step: 10, loss: 0.0038474781904369593
step: 20, loss: 0.009054141119122505
step: 30, loss: 0.003963193856179714
step: 40, loss: 0.001501589547842741
step: 50, loss: 0.0021325524430722
step: 60, loss: 0.029326820746064186
step: 70, loss: 0.00047153234481811523
step: 80, loss: 0.0009943245677277446
step: 90, loss: 0.0003466050839051604
step: 100, loss: 0.00025060930056497455
step: 110, loss: 0.0031481944024562836
step: 120, loss: 0.0011561147402971983
step: 130, loss: 0.07073356211185455
step: 140, loss: 0.00038281988236121833
step: 150, loss: 0.004731041844934225
step: 160, loss: 0.00044245546450838447
step: 170, loss: 0.00042380724335089326
step: 180, loss: 0.0006316456710919738
step: 190, loss: 0.056081924587488174
step: 200, loss: 0.006651486270129681
step: 210, loss: 0.0007996744243428111
step: 220, loss: 0.023867463693022728
step: 230, loss: 0.04478580504655838
/usr/local/lib/python3.7/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=0.9774774774774775, f1=0.9719416386083053,
best_f1=0.9744160177975528
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.02386412024497986
step: 10, loss: 0.16838829219341278
step: 20, loss: 0.0008366542169824243
step: 30, loss: 0.0003401228168513626
step: 40, loss: 0.000304636312648654
step: 50, loss: 0.003178281243890524
step: 60, loss: 0.0028689305763691664
step: 70, loss: 0.0008750408887863159
step: 80, loss: 0.002996403956785798
step: 90, loss: 0.000603526015765965
step: 100, loss: 0.0002298729814356193
step: 110, loss: 0.0002613450342323631
step: 120, loss: 0.00012187246466055512
step: 130, loss: 0.001259378157556057
step: 140, loss: 0.10375896841287613
step: 150, loss: 0.0011350007262080908
step: 160, loss: 0.004321396350860596
step: 170, loss: 0.005764688365161419
step: 180, loss: 0.002524986630305648
step: 190, loss: 0.0008744903025217354
step: 200, loss: 0.06150399148464203
step: 210, loss: 0.0006697108619846404
step: 220, loss: 0.0016305353492498398
step: 230, loss: 0.04945450648665428
/usr/local/lib/python3.7/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=0.9744160177975528, f1=0.9700332963374029,
best_f1=0.9744160177975528
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000379624922061339
step: 10, loss: 0.0008765202946960926
step: 20, loss: 0.001978089800104499
step: 30, loss: 0.002850067103281617
step: 40, loss: 0.00017128404579125345
step: 50, loss: 0.003163137473165989
step: 60, loss: 0.0002171387750422582
step: 70, loss: 0.0002191350213252008
step: 80, loss: 0.06496084481477737
step: 90, loss: 0.019212430343031883
step: 100, loss: 0.005915571469813585
step: 110, loss: 0.0001557495561428368
step: 120, loss: 0.05893145129084587
step: 130, loss: 0.0008768697734922171
step: 140, loss: 0.08047307282686234
step: 150, loss: 9.621905337553471e-05

```

```

step: 160, loss: 0.00025477376766502857
step: 170, loss: 0.00047644664300605655
step: 180, loss: 0.0006776999798603356
step: 190, loss: 0.00036173377884551883
step: 200, loss: 0.0005192013341002166
step: 210, loss: 9.835360833676532e-05
step: 220, loss: 0.02895740605890751
step: 230, loss: 0.003147685434669256
/usr/local/lib/python3.7/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=0.9742441209406495, f1=0.9720044792833147,
best_f1=0.9744160177975528
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00028225479763932526
step: 10, loss: 0.0009010586072690785
step: 20, loss: 0.00014891894534230232
step: 30, loss: 0.00017218473658431321
step: 40, loss: 6.574078724952415e-05
step: 50, loss: 0.00247390428557992
step: 60, loss: 0.07421936839818954
step: 70, loss: 0.000786582357250154
step: 80, loss: 0.00020473671611398458
step: 90, loss: 0.00015261011139955372
step: 100, loss: 8.003915718290955e-05
step: 110, loss: 9.968419908545911e-05
step: 120, loss: 0.008806821890175343
step: 130, loss: 0.00014763094077352434
step: 140, loss: 0.032244060188531876
step: 150, loss: 0.000343245716067031
step: 160, loss: 0.00024685979587957263
step: 170, loss: 0.0010008005192503333
step: 180, loss: 0.00023318931926041842
step: 190, loss: 0.0003615114255808294
step: 200, loss: 0.000325356493704021
step: 210, loss: 9.519107697997242e-05
step: 220, loss: 0.0009692957391962409
step: 230, loss: 0.00012843040167354047
/usr/local/lib/python3.7/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=0.976324689966178, f1=0.9738933030646991,
best_f1=0.9744160177975528
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 9.651399159338325e-05
step: 10, loss: 0.00024385255528613925
step: 20, loss: 5.351012805476785e-05
step: 30, loss: 0.0029047143179923296
step: 40, loss: 0.015272864140570164
step: 50, loss: 0.0020797005854547024
step: 60, loss: 0.0001941984664881602
step: 70, loss: 0.00015983385674189776
step: 80, loss: 0.00012700911611318588
step: 90, loss: 0.00031303352443501353
step: 100, loss: 0.000161807969561778
step: 110, loss: 0.00019152293680235744
step: 120, loss: 9.5920026069507e-05
step: 130, loss: 0.00020101950212847441
step: 140, loss: 5.1011516916332766e-05
step: 150, loss: 4.3211814045207575e-05
step: 160, loss: 6.068991933716461e-05
step: 170, loss: 0.0001462393265683204
step: 180, loss: 0.001554490183480084
step: 190, loss: 0.0002167365310015157
step: 200, loss: 8.766174141783267e-05
step: 210, loss: 4.961070590070449e-05
step: 220, loss: 0.0004805048811249435
step: 230, loss: 0.004242084454745054
/usr/local/lib/python3.7/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=0.9764309764309763, f1=0.9741863075196409,
best_f1=0.9744160177975528
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0005102892755530775
step: 10, loss: 0.00014082375855650753
step: 20, loss: 5.5157266615424305e-05
step: 30, loss: 0.0001707077753962949
step: 40, loss: 4.514832835411653e-05
step: 50, loss: 6.134330033091828e-05
step: 60, loss: 0.0018634628504514694
step: 70, loss: 5.320840864442289e-05
step: 80, loss: 6.38930723653175e-05
step: 90, loss: 4.929310307488777e-05
step: 100, loss: 0.00012376737140584737
step: 110, loss: 0.0015370417386293411
step: 120, loss: 0.0005075154476799071
step: 130, loss: 8.510463521815836e-05
step: 140, loss: 0.00019879642059095204
step: 150, loss: 0.00014501172699965537

```



```

step: 160, loss: 0.00038172677159309387
step: 170, loss: 8.149763016263023e-05
step: 180, loss: 7.232496136566624e-05
step: 190, loss: 0.000123962527140975
step: 200, loss: 0.006147574633359909
step: 210, loss: 7.390631071757525e-05
step: 220, loss: 0.04440746456384659
step: 230, loss: 7.920042844489217e-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=0.9774774774774775, f1=0.9751131221719457,
best_f1=0.9744160177975528
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00010120290971826762
step: 10, loss: 6.424175080610439e-05
step: 20, loss: 9.047716594068334e-05
step: 30, loss: 0.0011068364838138223
step: 40, loss: 0.0001338272704742849
step: 50, loss: 8.108533802442253e-05
step: 60, loss: 9.606603998690844e-05
step: 70, loss: 7.476173777831718e-05
step: 80, loss: 5.599495125352405e-05
step: 90, loss: 4.7345954953925684e-05
step: 100, loss: 8.358025661436841e-05
step: 110, loss: 6.569801189471036e-05
step: 120, loss: 0.00018434794037602842
step: 130, loss: 0.00015215018356684595
step: 140, loss: 0.003705605398863554
step: 150, loss: 0.029534442350268364
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 160, loss: 0.04490635171532631
step: 170, loss: 6.885377661092207e-05
step: 180, loss: 0.00022120343055576086
step: 190, loss: 0.0005763346562162042
step: 200, loss: 7.669049955438823e-05
step: 210, loss: 8.117772813420743e-05
step: 220, loss: 8.219719165936112e-05
step: 230, loss: 5.0221824494656175e-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=0.9765363128491621, f1=0.9776286353467561,
best_f1=0.9744160177975528
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 5.8805344451684505e-05
step: 10, loss: 3.786908200709149e-05
step: 20, loss: 0.00027935142861679196
step: 30, loss: 6.486464553745463e-05
step: 40, loss: 4.032070137327537e-05
step: 50, loss: 0.0008448283770121634
step: 60, loss: 4.5750242861686274e-05
step: 70, loss: 5.2013503591297194e-05
step: 80, loss: 7.388518861262128e-05
step: 90, loss: 6.849558849353343e-05
step: 100, loss: 0.003154020057991147
step: 110, loss: 0.00010424799256725237
step: 120, loss: 5.573014641413465e-05
step: 130, loss: 7.642833952559158e-05
step: 140, loss: 5.494633660418913e-05
step: 150, loss: 7.340431329794228e-05
step: 160, loss: 3.8495418266393244e-05
step: 170, loss: 7.127254502847791e-05
step: 180, loss: 7.339869625866413e-05
step: 190, loss: 8.422934479312971e-05
step: 200, loss: 5.5275566410273314e-05
step: 210, loss: 0.00028418947476893663
step: 220, loss: 8.202960452763364e-05
step: 230, loss: 3.320651376270689e-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=0.9776286353467561, f1=0.9764837625979844,
best_f1=0.9744160177975528
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 4.565480412566103e-05
step: 10, loss: 8.505582081852481e-05
step: 20, loss: 4.899253326584585e-05
step: 30, loss: 0.0001248875487362966
step: 40, loss: 6.626794493058696e-05
step: 50, loss: 6.635022873524576e-05
step: 60, loss: 4.2831354221561924e-05
step: 70, loss: 5.80657142563723e-05
step: 80, loss: 0.0001760348677635193
step: 90, loss: 3.3168929803650826e-05
step: 100, loss: 7.066989928716794e-05
step: 110, loss: 5.85796260565985e-05
step: 120, loss: 6.5156928030774e-05
step: 130, loss: 7.300466677406803e-05
step: 140, loss: 0.0011246141511946917

```

```

step: 150, loss: 6.747177394572645e-05
step: 160, loss: 5.6266017054440454e-05
step: 170, loss: 4.62211792182643e-05
step: 180, loss: 4.878403342445381e-05
step: 190, loss: 0.0008474108763039112
step: 200, loss: 4.324851761339232e-05
step: 210, loss: 0.04010411351919174
step: 220, loss: 0.00029012374579906464
step: 230, loss: 8.210210944525898e-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=0.9775784753363228, f1=0.9753363228699552,
best_f1=0.9744160177975528
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
2473it [00:07, 319.33it/s]
load_f1 = 0.9730941704035874
real_f1 = 0.9730337078651685
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-

```

```
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:11, 371.84it/s]
```

3.3.2 DBLP-GoogleScholar - Running the matcher

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

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

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

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

Defaults for this optimization level are:

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

```

keep_batchnorm_fp32      : True
master_weights           : True
loss_scale               : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                   : True
opt_level                 : 02
cast_model_type           : torch.float16
patch_torch_functions     : False
keep_batchnorm_fp32      : True
master_weights           : True
loss_scale               : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.7947001457214355
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.46695882081985474
step: 20, loss: 0.5028555989265442
step: 30, loss: 0.42597922682762146
step: 40, loss: 0.3084474503993988
step: 50, loss: 0.2863752841949463
step: 60, loss: 0.294549822807312
step: 70, loss: 0.1970454603433609
step: 80, loss: 0.056779008358716965
step: 90, loss: 0.05835087597370148
step: 100, loss: 0.23125214874744415
step: 110, loss: 0.12812276184558868
step: 120, loss: 0.03310488536953926
step: 130, loss: 0.01604257896542549
step: 140, loss: 0.12865841388702393
step: 150, loss: 0.03751572221517563
step: 160, loss: 0.23658594489097595
step: 170, loss: 0.24785906076431274
step: 180, loss: 0.12423828989267349

```

```

step: 190, loss: 0.12849277257919312
step: 200, loss: 0.1502254754304886
step: 210, loss: 0.12587156891822815
step: 220, loss: 0.10080358386039734
step: 230, loss: 0.11809763312339783
step: 240, loss: 0.08276166021823883
step: 250, loss: 0.09134896844625473
step: 260, loss: 0.09027614444494247
step: 270, loss: 0.020283440127968788
step: 280, loss: 0.1137036606669426
step: 290, loss: 0.055300816893577576
step: 300, loss: 0.25423508882522583
step: 310, loss: 0.1097802072763443
step: 320, loss: 0.03380747139453888
step: 330, loss: 0.13663770258426666
step: 340, loss: 0.11564838886260986
step: 350, loss: 0.11186838895082474
step: 360, loss: 0.05155062675476074
step: 370, loss: 0.1249004378914833
step: 380, loss: 0.19728711247444153
step: 390, loss: 0.056735213845968246
step: 400, loss: 0.015508783981204033
step: 410, loss: 0.0503629595041275
step: 420, loss: 0.019393105059862137
step: 430, loss: 0.14385195076465607
step: 440, loss: 0.05621829628944397
step: 450, loss: 0.021010689437389374
step: 460, loss: 0.28062012791633606
step: 470, loss: 0.25211524963378906
step: 480, loss: 0.2703983187675476
step: 490, loss: 0.05122102051973343
step: 500, loss: 0.020690523087978363
step: 510, loss: 0.03520850092172623
step: 520, loss: 0.11154138296842575
step: 530, loss: 0.06054835021495819
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9187935034802784, f1=0.9180934752429432,
best_f1=0.9180934752429432
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0959945023059845

```

step: 10, loss: 0.1493905782699585
step: 20, loss: 0.19151616096496582
step: 30, loss: 0.13843484222888947
step: 40, loss: 0.013197964988648891
step: 50, loss: 0.009411790408194065
step: 60, loss: 0.10025659203529358
step: 70, loss: 0.24231359362602234
step: 80, loss: 0.021523764356970787
step: 90, loss: 0.0025354810059070587
step: 100, loss: 0.21247908473014832
step: 110, loss: 0.06022059544920921
step: 120, loss: 0.08538565039634705
step: 130, loss: 0.021048635244369507
step: 140, loss: 0.04957534000277519
step: 150, loss: 0.13333283364772797
step: 160, loss: 0.01660328544676304
step: 170, loss: 0.2753528654575348
step: 180, loss: 0.003944186493754387
step: 190, loss: 0.013961318880319595
step: 200, loss: 0.02874085307121277
step: 210, loss: 0.03627195581793785
step: 220, loss: 0.14932626485824585
step: 230, loss: 0.029394835233688354
step: 240, loss: 0.0537639744579792
step: 250, loss: 0.02914741449058056
step: 260, loss: 0.005674588028341532
step: 270, loss: 0.21939507126808167
step: 280, loss: 0.0519149973988533
step: 290, loss: 0.08756396919488907
step: 300, loss: 0.012860656715929508
step: 310, loss: 0.06536581367254257
step: 320, loss: 0.22735823690891266
step: 330, loss: 0.01212403830140829
step: 340, loss: 0.006435676943510771
step: 350, loss: 0.038229916244745255
step: 360, loss: 0.01500582043081522
step: 370, loss: 0.026158099994063377
step: 380, loss: 0.050828222185373306
step: 390, loss: 0.02082745172083378
step: 400, loss: 0.0530865415930748
step: 410, loss: 0.0038185445591807365
step: 420, loss: 0.03360790014266968
step: 430, loss: 0.02016773819923401
step: 440, loss: 0.007650554180145264
step: 450, loss: 0.024901889264583588
step: 460, loss: 0.2761981785297394
step: 470, loss: 0.08137054741382599
step: 480, loss: 0.2179899960756302

```

step: 490, loss: 0.040677547454833984
step: 500, loss: 0.04458296671509743
step: 510, loss: 0.05080220475792885
step: 520, loss: 0.08789418637752533
step: 530, loss: 0.18402214348316193
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9280373831775701, f1=0.9288040949278734,
best_f1=0.9288040949278734
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009550523944199085
step: 10, loss: 0.08555442839860916
step: 20, loss: 0.16015511751174927
step: 30, loss: 0.20070558786392212
step: 40, loss: 0.004149328451603651
step: 50, loss: 0.018443945795297623
step: 60, loss: 0.002447259845212102
step: 70, loss: 0.005630738101899624
step: 80, loss: 0.011540373787283897
step: 90, loss: 0.1516328901052475
step: 100, loss: 0.0859374850988388
step: 110, loss: 0.014873504638671875
step: 120, loss: 0.010736743919551373
step: 130, loss: 0.010726315900683403
step: 140, loss: 0.009572623297572136
step: 150, loss: 0.0030374929774552584
step: 160, loss: 0.0035306462086737156
step: 170, loss: 0.015476738102734089
step: 180, loss: 0.03156689181923866
step: 190, loss: 0.018505776301026344
step: 200, loss: 0.031632501631975174
step: 210, loss: 0.015100432559847832
step: 220, loss: 0.013567337766289711
step: 230, loss: 0.02577531337738037
step: 240, loss: 0.02873336896300316
step: 250, loss: 0.009202074259519577
step: 260, loss: 0.00694301538169384
step: 270, loss: 0.001869776169769466
step: 280, loss: 0.046972014009952545
step: 290, loss: 0.06752537190914154
step: 300, loss: 0.13924026489257812

```



```

step: 310, loss: 0.08601494878530502
step: 320, loss: 0.18625813722610474
step: 330, loss: 0.00313957454636693
step: 340, loss: 0.020479660481214523
step: 350, loss: 0.0036630660761147738
step: 360, loss: 0.01582931913435459
step: 370, loss: 0.0882115289568901
step: 380, loss: 0.06898751854896545
step: 390, loss: 0.014799683354794979
step: 400, loss: 0.06848881393671036
step: 410, loss: 0.009501241147518158
step: 420, loss: 0.09274307638406754
step: 430, loss: 0.05645501986145973
step: 440, loss: 0.029741309583187103
step: 450, loss: 0.00769137404859066
step: 460, loss: 0.051795843988657
step: 470, loss: 0.004982156679034233
step: 480, loss: 0.0024258745834231377
step: 490, loss: 0.011972843669354916
step: 500, loss: 0.08487039059400558
step: 510, loss: 0.001451462390832603
step: 520, loss: 0.001846817322075367
step: 530, loss: 0.038404837250709534
/usr/local/lib/python3.7/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=0.9272976680384087, f1=0.9240681086056144,
best_f1=0.9288040949278734
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0020074977073818445
step: 10, loss: 0.006827730685472488
step: 20, loss: 0.03136308118700981
step: 30, loss: 0.013623111881315708
step: 40, loss: 0.011205771937966347
step: 50, loss: 0.005793888121843338
step: 60, loss: 0.0015216395258903503
step: 70, loss: 0.004747115075588226
step: 80, loss: 0.03499254211783409
step: 90, loss: 0.05539379268884659
step: 100, loss: 0.023147935047745705
step: 110, loss: 0.003728077979758382
step: 120, loss: 0.04174205660820007
step: 130, loss: 0.08200659602880478
step: 140, loss: 0.027065709233283997
step: 150, loss: 0.0018284769030287862
step: 160, loss: 0.032800912857055664

```

```

step: 170, loss: 0.025245487689971924
step: 180, loss: 0.0032476799096912146
step: 190, loss: 0.0016109615098685026
step: 200, loss: 0.002176244044676423
step: 210, loss: 0.003683411283418536
step: 220, loss: 0.001381036126986146
step: 230, loss: 0.2225790172815323
step: 240, loss: 0.003185942303389311
step: 250, loss: 0.19082903861999512
step: 260, loss: 0.12418344616889954
step: 270, loss: 0.05587548762559891
step: 280, loss: 0.006023733876645565
step: 290, loss: 0.1525019407272339
step: 300, loss: 0.037953585386276245
step: 310, loss: 0.008695810101926327
step: 320, loss: 0.022620409727096558
step: 330, loss: 0.09685304760932922
step: 340, loss: 0.002293191384524107
step: 350, loss: 0.0006296906503848732
step: 360, loss: 0.012926402501761913
step: 370, loss: 0.05006836727261543
step: 380, loss: 0.0020888629369437695
step: 390, loss: 0.014231636188924313
step: 400, loss: 0.009218864142894745
step: 410, loss: 0.01422074344009161
step: 420, loss: 0.019987544044852257
step: 430, loss: 0.01063730288296938
step: 440, loss: 0.06934060156345367
step: 450, loss: 0.002708078594878316
step: 460, loss: 0.005425776354968548
step: 470, loss: 0.018759451806545258
step: 480, loss: 0.0038755894638597965
step: 490, loss: 0.03662888705730438
step: 500, loss: 0.0008496090304106474
step: 510, loss: 0.010135776363313198
step: 520, loss: 0.0053984071128070354
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 530, loss: 0.07703624665737152
/usr/local/lib/python3.7/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=0.9266266728195662, f1=0.9319129226493746,
best_f1=0.9288040949278734
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003915663808584213
step: 10, loss: 0.030616765841841698

```

step: 20, loss: 0.004962734412401915
step: 30, loss: 0.054525382816791534
step: 40, loss: 0.016759559512138367
step: 50, loss: 0.0030737188644707203
step: 60, loss: 0.0015648632543161511
step: 70, loss: 0.0022589287254959345
step: 80, loss: 0.0009522387990728021
step: 90, loss: 0.008466709405183792
step: 100, loss: 0.0035556333605200052
step: 110, loss: 0.0013962042285129428
step: 120, loss: 0.002593263052403927
step: 130, loss: 0.0014221223536878824
step: 140, loss: 0.00026336469454690814
step: 150, loss: 0.001027016551233828
step: 160, loss: 0.004398945719003677
step: 170, loss: 0.01573798432946205
step: 180, loss: 0.0003898813738487661
step: 190, loss: 0.0004025848757009953
step: 200, loss: 0.0007434449507854879
step: 210, loss: 0.0017468997975811362
step: 220, loss: 0.00019584590336307883
step: 230, loss: 0.000550821830984205
step: 240, loss: 0.0728766918182373
step: 250, loss: 0.0010922309011220932
step: 260, loss: 0.002778031397610903
step: 270, loss: 0.01367843896150589
step: 280, loss: 0.03401636704802513
step: 290, loss: 0.00487834308296442
step: 300, loss: 0.0808127224445343
step: 310, loss: 0.0009008661727420986
step: 320, loss: 0.011562509462237358
step: 330, loss: 0.004785096738487482
step: 340, loss: 0.010250451043248177
step: 350, loss: 0.014758610166609287
step: 360, loss: 0.0007697077817283571
step: 370, loss: 0.0007612006738781929
step: 380, loss: 0.025535909458994865
step: 390, loss: 0.003975275903940201
step: 400, loss: 0.019307779148221016
step: 410, loss: 0.0003525727370288223
step: 420, loss: 0.01876913011074066
step: 430, loss: 0.012606148608028889
step: 440, loss: 0.03529200330376625
step: 450, loss: 0.0007342698518186808
step: 460, loss: 0.0018789247842505574
step: 470, loss: 0.001598559902049601
step: 480, loss: 0.04269682615995407
step: 490, loss: 0.00246114912442863

```

step: 500, loss: 0.0036055641248822212
step: 510, loss: 0.0009467257186770439
step: 520, loss: 0.00031991966534405947
step: 530, loss: 0.003146413015201688
/usr/local/lib/python3.7/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=0.9234972677595629, f1=0.9230064161319891,
best_f1=0.9288040949278734
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006285231793299317
step: 10, loss: 0.00851090531796217
step: 20, loss: 0.0018402975983917713
step: 30, loss: 0.0071799797005951405
step: 40, loss: 0.0009650353458710015
step: 50, loss: 0.03244636952877045
step: 60, loss: 0.0037183111999183893
step: 70, loss: 0.04530522599816322
step: 80, loss: 0.0015543497866019607
step: 90, loss: 0.00035866929101757705
step: 100, loss: 0.0028540154453366995
step: 110, loss: 0.0024823863059282303
step: 120, loss: 0.001458123093470931
step: 130, loss: 0.0018256563926115632
step: 140, loss: 0.004640019498765469
step: 150, loss: 0.00018941200687550008
step: 160, loss: 0.00011985543096670881
step: 170, loss: 0.00010798019502544776
step: 180, loss: 0.0031915351282805204
step: 190, loss: 0.010748863220214844
step: 200, loss: 0.00036384782288223505
step: 210, loss: 0.00490452628582716
step: 220, loss: 0.00023506241268478334
step: 230, loss: 0.00013443766511045396
step: 240, loss: 0.0002574326645117253
step: 250, loss: 0.00017410186410415918
step: 260, loss: 0.0008088296162895858
step: 270, loss: 0.0005261480691842735
step: 280, loss: 0.0015932448441162705
step: 290, loss: 0.0008118769619613886
step: 300, loss: 0.0012003767769783735
step: 310, loss: 0.00030805732239969075
step: 320, loss: 0.08629534393548965
step: 330, loss: 0.0009687934070825577
step: 340, loss: 0.0004949131398461759
step: 350, loss: 0.04177750647068024

```

```

step: 360, loss: 0.0072936504147946835
step: 370, loss: 0.002644418040290475
step: 380, loss: 0.005576377268880606
step: 390, loss: 0.018978888168931007
step: 400, loss: 0.00031649201991967857
step: 410, loss: 0.0003800092963501811
step: 420, loss: 0.02153507061302662
step: 430, loss: 0.00016809852968435735
step: 440, loss: 0.036422234028577805
step: 450, loss: 0.0017543506110087037
step: 460, loss: 0.0015133833512663841
step: 470, loss: 0.03581291437149048
step: 480, loss: 0.014947904273867607
step: 490, loss: 8.926400187192485e-05
step: 500, loss: 0.01710793934762478
step: 510, loss: 0.0001130078817368485
step: 520, loss: 0.0003697130596265197
step: 530, loss: 0.005753600969910622
/usr/local/lib/python3.7/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=0.9259431765253843, f1=0.9290382819794585,
best_f1=0.9288040949278734
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.08538710325956345
step: 10, loss: 0.0007938421913422644
step: 20, loss: 0.005487389396876097
step: 30, loss: 0.027257997542619705
step: 40, loss: 8.935458754422143e-05
step: 50, loss: 0.00031890536774881184
step: 60, loss: 0.01187078095972538
step: 70, loss: 0.00035964095150120556
step: 80, loss: 0.006713251583278179
step: 90, loss: 0.00023603622685186565
step: 100, loss: 0.003916262648999691
step: 110, loss: 0.06805774569511414
step: 120, loss: 0.00010689444025047123
step: 130, loss: 0.0048807477578520775
step: 140, loss: 0.02580741047859192
step: 150, loss: 0.0003446304763201624
step: 160, loss: 0.09922707080841064
step: 170, loss: 0.006633567623794079
step: 180, loss: 0.0003517276491038501
step: 190, loss: 0.00016841065371409059
step: 200, loss: 0.0219923947006464
step: 210, loss: 0.0003202518855687231

```

```

step: 220, loss: 0.00012459393474273384
step: 230, loss: 0.0009714693296700716
step: 240, loss: 0.0010762307792901993
step: 250, loss: 0.0005039707757532597
step: 260, loss: 0.010187765583395958
step: 270, loss: 0.0006624360103160143
step: 280, loss: 0.004148669075220823
step: 290, loss: 0.002624969929456711
step: 300, loss: 0.0007785825291648507
step: 310, loss: 0.00021764490520581603
step: 320, loss: 0.017724407836794853
step: 330, loss: 0.0007015076116658747
step: 340, loss: 0.0005847770371474326
step: 350, loss: 0.00012468760542105883
step: 360, loss: 0.000877136189956218
step: 370, loss: 0.001329951104708016
step: 380, loss: 0.0002939462137874216
step: 390, loss: 9.227586269844323e-05
step: 400, loss: 0.0037839903961867094
step: 410, loss: 0.0006970927934162319
step: 420, loss: 0.00015608480316586792
step: 430, loss: 6.255619518924505e-05
step: 440, loss: 0.00011280965554760769
step: 450, loss: 0.027982400730252266
step: 460, loss: 0.04932491108775139
step: 470, loss: 0.009115401655435562
step: 480, loss: 0.0005791953299194574
step: 490, loss: 0.0001497576740803197
step: 500, loss: 0.0005413030739873648
step: 510, loss: 0.0002177378482883796
step: 520, loss: 0.011479602195322514
step: 530, loss: 0.0002959244011435658
/usr/local/lib/python3.7/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=0.92524682651622, f1=0.9210526315789475,
best_f1=0.9288040949278734
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.000593745440710336
step: 10, loss: 0.0004717808333225548
step: 20, loss: 0.03064730390906334
step: 30, loss: 0.0011432478204369545
step: 40, loss: 0.00018938936409540474
step: 50, loss: 0.0010015321895480156
step: 60, loss: 0.0016768226632848382
step: 70, loss: 0.00017377127369400114

```

step: 80, loss: 0.0001278519630432129
step: 90, loss: 0.00032833218574523926
step: 100, loss: 0.00039071368519216776
step: 110, loss: 0.05778283625841141
step: 120, loss: 0.01536567322909832
step: 130, loss: 0.0003209684800822288
step: 140, loss: 0.0001075027757906355
step: 150, loss: 0.0013822378823533654
step: 160, loss: 0.00383303826674819
step: 170, loss: 0.00022242768318392336
step: 180, loss: 0.0008021413814276457
step: 190, loss: 0.0003451644442975521
step: 200, loss: 0.001138058491051197
step: 210, loss: 0.0004797483852598816
step: 220, loss: 0.026161447167396545
step: 230, loss: 0.0011522063286975026
step: 240, loss: 4.317398270359263e-05
step: 250, loss: 0.00018479528080206364
step: 260, loss: 0.0002289659605594352
step: 270, loss: 8.480909309582785e-05
step: 280, loss: 8.68153729243204e-05
step: 290, loss: 0.005260927602648735
step: 300, loss: 7.702036964474246e-05
step: 310, loss: 0.010688574984669685
step: 320, loss: 4.887668910669163e-05
step: 330, loss: 0.017540492117404938
step: 340, loss: 0.002471037907525897
step: 350, loss: 0.0031540067866444588
step: 360, loss: 5.257542579784058e-05
step: 370, loss: 0.00036002337583340704
step: 380, loss: 0.0009573809220455587
step: 390, loss: 7.281605940079316e-05
step: 400, loss: 0.0347508080303669
step: 410, loss: 0.00027378470986150205
step: 420, loss: 0.0013697617687284946
step: 430, loss: 9.558689635014161e-05
step: 440, loss: 0.00011994943633908406
step: 450, loss: 0.00010260627459501848
step: 460, loss: 0.0004451481800060719
step: 470, loss: 0.0039563290774822235
step: 480, loss: 7.458475010935217e-05
step: 490, loss: 7.078804628690705e-05
step: 500, loss: 0.00309460680000484
step: 510, loss: 0.00020937222870998085
step: 520, loss: 0.0031924843788146973
step: 530, loss: 0.0001980937086045742
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9308411214953269, f1=0.9248014946286782,
best_f1=0.9248014946286782
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00041321825119666755
step: 10, loss: 0.0001532991445856169
step: 20, loss: 0.00032470765290781856
step: 30, loss: 0.00010329476208426058
step: 40, loss: 0.00046160456258803606
step: 50, loss: 6.990874680923298e-05
step: 60, loss: 7.64461001381278e-05
step: 70, loss: 0.0947294533252716
step: 80, loss: 6.919458246557042e-05
step: 90, loss: 0.0006031045340932906
step: 100, loss: 6.043163375579752e-05
step: 110, loss: 0.022636571899056435
step: 120, loss: 0.0006032632663846016
step: 130, loss: 0.011728262528777122
step: 140, loss: 0.0006776417139917612
step: 150, loss: 3.4353310184087604e-05
step: 160, loss: 4.3506715883268043e-05
step: 170, loss: 3.548588210833259e-05
step: 180, loss: 3.7074871215736493e-05
step: 190, loss: 5.3420364565681666e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 200, loss: 0.002906601410359144
step: 210, loss: 3.805948290391825e-05
step: 220, loss: 8.134308882290497e-05
step: 230, loss: 4.08639753004536e-05
step: 240, loss: 7.917475159047171e-05
step: 250, loss: 0.0010367922950536013
step: 260, loss: 0.15346771478652954
step: 270, loss: 0.0003748357994481921
step: 280, loss: 4.273786544217728e-05
step: 290, loss: 0.0001184112552437
step: 300, loss: 0.00012261734809726477
step: 310, loss: 4.238643668941222e-05
step: 320, loss: 0.0001781048340490088
step: 330, loss: 0.0333373062312603
step: 340, loss: 4.214563159621321e-05
step: 350, loss: 4.3818599806400016e-05
step: 360, loss: 0.008819849230349064

```



```

step: 370, loss: 0.00014495216601062566
step: 380, loss: 3.0863157007843256e-05
step: 390, loss: 3.336638110340573e-05
step: 400, loss: 0.0015834305668249726
step: 410, loss: 0.00018911270308308303
step: 420, loss: 0.00021176078007556498
step: 430, loss: 0.000169418373843655
step: 440, loss: 0.00020000609220005572
step: 450, loss: 0.00020719121675938368
step: 460, loss: 0.015307779423892498
step: 470, loss: 6.30874274065718e-05
step: 480, loss: 5.6756969570415094e-05
step: 490, loss: 0.001386580290272832
step: 500, loss: 0.00014195429685059935
step: 510, loss: 0.00013815292913932353
step: 520, loss: 2.8601756639545783e-05
step: 530, loss: 0.00019904137297999114
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9318497913769124, f1=0.9285051067780873,
best_f1=0.9285051067780873
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003142348723486066
step: 10, loss: 0.0004869860131293535
step: 20, loss: 0.00015162545605562627
step: 30, loss: 3.104942516074516e-05
step: 40, loss: 4.69245424028486e-05
step: 50, loss: 0.0004906348767690361
step: 60, loss: 4.856373925576918e-05
step: 70, loss: 0.00011244878987781703
step: 80, loss: 0.0011217871215194464
step: 90, loss: 4.317663115216419e-05
step: 100, loss: 5.610206426354125e-05
step: 110, loss: 0.0009641089709475636
step: 120, loss: 4.230990452924743e-05
step: 130, loss: 9.094714187085629e-05
step: 140, loss: 5.713423524866812e-05
step: 150, loss: 4.510459257289767e-05
step: 160, loss: 7.309905049623922e-05
step: 170, loss: 0.0010903176153078675
step: 180, loss: 0.0004296725383028388

```

```

step: 190, loss: 3.0475843232125044e-05
step: 200, loss: 5.6232827773783356e-05
step: 210, loss: 7.002988422755152e-05
step: 220, loss: 3.5888013371732086e-05
step: 230, loss: 4.8266429075738415e-05
step: 240, loss: 8.586351032136008e-05
step: 250, loss: 0.00021311129967216402
step: 260, loss: 0.0004301642766222358
step: 270, loss: 2.1636114979628474e-05
step: 280, loss: 8.729590626899153e-05
step: 290, loss: 0.00018650996207725257
step: 300, loss: 0.01603311486542225
step: 310, loss: 0.00026217009872198105
step: 320, loss: 0.00023530710313934833
step: 330, loss: 0.0004275463579688221
step: 340, loss: 0.0005502037820406258
step: 350, loss: 4.161382094025612e-05
step: 360, loss: 3.069195736316033e-05
step: 370, loss: 0.0012012096121907234
step: 380, loss: 4.7253714001271874e-05
step: 390, loss: 4.464848825591616e-05
step: 400, loss: 0.00015802467532921582
step: 410, loss: 0.0005455639329738915
step: 420, loss: 0.0008327779942192137
step: 430, loss: 3.245373227400705e-05
step: 440, loss: 3.7191704905126244e-05
step: 450, loss: 0.00029100169194862247
step: 460, loss: 0.0003925594501197338
step: 470, loss: 3.5508139262674376e-05
step: 480, loss: 5.886880535399541e-05
step: 490, loss: 0.040175702422857285
step: 500, loss: 0.000244376715272665
step: 510, loss: 0.0010041446657851338
step: 520, loss: 7.861879566917196e-05
step: 530, loss: 0.001626845565624535
/usr/local/lib/python3.7/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=0.9299303944315545, f1=0.9310504396112911,
best_f1=0.9285051067780873
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00026077224174514413
step: 10, loss: 0.0007174288621172309
step: 20, loss: 3.502919935272075e-05
step: 30, loss: 4.936387267662212e-05
step: 40, loss: 0.00010392726107966155

```

step: 50, loss: 0.00035513253533281386
step: 60, loss: 2.9711411116295494e-05
step: 70, loss: 2.8035474315402098e-05
step: 80, loss: 5.811798109789379e-05
step: 90, loss: 4.860871558776125e-05
step: 100, loss: 4.877640458289534e-05
step: 110, loss: 5.957161556580104e-05
step: 120, loss: 0.0002105984603986144
step: 130, loss: 0.00017582204600330442
step: 140, loss: 0.0003023489553015679
step: 150, loss: 0.0003214891185052693
step: 160, loss: 0.0036213346756994724
step: 170, loss: 0.0001974810875253752
step: 180, loss: 5.4453423217637464e-05
step: 190, loss: 0.00016876951849553734
step: 200, loss: 5.663774936692789e-05
step: 210, loss: 0.000712823064532131
step: 220, loss: 0.0001931280130520463
step: 230, loss: 0.0028600571677088737
step: 240, loss: 3.653595194919035e-05
step: 250, loss: 0.0001405440125381574
step: 260, loss: 6.42359591438435e-05
step: 270, loss: 0.0016684734728187323
step: 280, loss: 0.0031840363517403603
step: 290, loss: 0.006338675040751696
step: 300, loss: 0.0024926115293055773
step: 310, loss: 0.002527299802750349
step: 320, loss: 0.0019518985645845532
step: 330, loss: 0.06111430749297142
step: 340, loss: 7.650638144696131e-05
step: 350, loss: 0.10783486068248749
step: 360, loss: 0.0002265557850478217
step: 370, loss: 0.0007138950750231743
step: 380, loss: 0.0018050186336040497
step: 390, loss: 0.12063372135162354
step: 400, loss: 3.398396438569762e-05
step: 410, loss: 0.0005841750535182655
step: 420, loss: 0.0004122567770536989
step: 430, loss: 0.00013039571058470756
step: 440, loss: 0.00011576511315070093
step: 450, loss: 0.00016634601342957467
step: 460, loss: 0.005256973672658205
step: 470, loss: 0.0002357935009058565
step: 480, loss: 4.4473734305938706e-05
step: 490, loss: 0.002832692349329591
step: 500, loss: 0.0003217712219338864
step: 510, loss: 3.541872138157487e-05
step: 520, loss: 2.079794285236858e-05

```

Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 530, loss: 2.654930176504422e-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=0.9245810055865922, f1=0.9301675977653632,
best_f1=0.9285051067780873
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00021277772611938417
step: 10, loss: 5.652465915773064e-05
step: 20, loss: 1.9184863049304113e-05
step: 30, loss: 2.8054431822965853e-05
step: 40, loss: 0.0002822953974828124
step: 50, loss: 0.001414341269992292
step: 60, loss: 2.6903251637122594e-05
step: 70, loss: 8.978566620498896e-05
step: 80, loss: 0.00018106828792952
step: 90, loss: 0.0001448599505238235
step: 100, loss: 0.05143105983734131
step: 110, loss: 6.759772077202797e-05
step: 120, loss: 0.0003390454512555152
step: 130, loss: 4.487484693527222e-05
step: 140, loss: 0.0001596478105057031
step: 150, loss: 0.00043172776349820197
step: 160, loss: 8.01050482550636e-05
step: 170, loss: 5.6038552429527044e-05
step: 180, loss: 0.00026459689252078533
step: 190, loss: 2.5942197680706158e-05
step: 200, loss: 0.000105927130789496
step: 210, loss: 8.557324326829985e-05
step: 220, loss: 3.134373764623888e-05
step: 230, loss: 0.00012694181350525469
step: 240, loss: 2.6012952730525285e-05
step: 250, loss: 0.00034640281228348613
step: 260, loss: 0.0009090694948099554
step: 270, loss: 0.00015624426305294037
step: 280, loss: 8.017166692297906e-05
step: 290, loss: 0.0005143460584804416
step: 300, loss: 0.0003932502877432853
step: 310, loss: 0.014078722335398197
step: 320, loss: 7.305465260287747e-05
step: 330, loss: 2.9313576305867173e-05
step: 340, loss: 0.0005295221344567835
step: 350, loss: 0.014440927654504776
step: 360, loss: 0.0035825653467327356
step: 370, loss: 3.394015948288143e-05

```

```

step: 380, loss: 3.174971425323747e-05
step: 390, loss: 3.466525231488049e-05
step: 400, loss: 8.003941184142604e-05
step: 410, loss: 0.00019187624275218695
step: 420, loss: 0.00135571020655334
step: 430, loss: 6.795836088713259e-05
step: 440, loss: 0.0014342055656015873
step: 450, loss: 0.0003906073106918484
step: 460, loss: 0.002818057080730796
step: 470, loss: 0.00018573709530755877
step: 480, loss: 0.0028722600545734167
step: 490, loss: 0.00019369993242435157
step: 500, loss: 0.0001675639650784433
step: 510, loss: 4.7624955186620355e-05
step: 520, loss: 0.002605760470032692
step: 530, loss: 0.00018968818767461926
/usr/local/lib/python3.7/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=0.9256047466910087, f1=0.926605504587156,
best_f1=0.9285051067780873
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009878583950921893
step: 10, loss: 4.960417572874576e-05
step: 20, loss: 0.00019517168402671814
step: 30, loss: 0.00016222054546233267
step: 40, loss: 0.00013516719627659768
step: 50, loss: 4.9611473514232785e-05
step: 60, loss: 4.277428161003627e-05
step: 70, loss: 6.599914195248857e-05
step: 80, loss: 0.0002894532517530024
step: 90, loss: 4.513545718509704e-05
step: 100, loss: 1.7065311112673953e-05
step: 110, loss: 0.0014442058745771646
step: 120, loss: 0.00011407010606490076
step: 130, loss: 5.6093878811225295e-05
step: 140, loss: 3.681518865050748e-05
step: 150, loss: 6.0892984038218856e-05
step: 160, loss: 0.00011836538760690019
step: 170, loss: 3.737434963113628e-05
step: 180, loss: 2.1412575733847916e-05
step: 190, loss: 7.682482828386128e-05
step: 200, loss: 0.0021874536760151386
step: 210, loss: 0.00010074173042085022
step: 220, loss: 2.3982844140846282e-05
step: 230, loss: 2.7413450879976153e-05

```

```

step: 240, loss: 6.12725198152475e-05
step: 250, loss: 0.01196235604584217
step: 260, loss: 0.0003852745285257697
step: 270, loss: 3.319375537103042e-05
step: 280, loss: 0.0007330093649215996
step: 290, loss: 3.5872530133929104e-05
step: 300, loss: 0.00010383569315308705
step: 310, loss: 8.067039743764326e-05
step: 320, loss: 4.3646188714774325e-05
step: 330, loss: 0.0002484005526639521
step: 340, loss: 2.406859493930824e-05
step: 350, loss: 0.0017444086261093616
step: 360, loss: 3.466262569418177e-05
step: 370, loss: 2.5647734219091944e-05
step: 380, loss: 7.267723412951455e-05
step: 390, loss: 3.9816932257963344e-05
step: 400, loss: 0.00036655389703810215
step: 410, loss: 3.233047027606517e-05
step: 420, loss: 4.6684763219673187e-05
step: 430, loss: 0.0021840224508196115
step: 440, loss: 0.0006285380804911256
step: 450, loss: 4.656887176679447e-05
step: 460, loss: 3.5834524169331416e-05
step: 470, loss: 3.218456913600676e-05
step: 480, loss: 0.0021458431147038937
step: 490, loss: 5.2368086471688e-05
step: 500, loss: 0.007198401261121035
step: 510, loss: 3.6169552913634107e-05
step: 520, loss: 5.4492298659170046e-05
step: 530, loss: 3.457489947322756e-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=0.9233698238933842, f1=0.9235880398671096,
best_f1=0.9285051067780873
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 3.211033617844805e-05
step: 10, loss: 0.000764021067880094
step: 20, loss: 1.919232636282686e-05
step: 30, loss: 3.831625508610159e-05
step: 40, loss: 2.7107602363685146e-05
step: 50, loss: 0.00017134100198745728
step: 60, loss: 5.191386298974976e-05
step: 70, loss: 1.826477819122374e-05
step: 80, loss: 0.00014112306234892458
step: 90, loss: 1.746015368553344e-05

```

```
step: 100, loss: 2.3055348719935864e-05
step: 110, loss: 6.620791100431234e-05
step: 120, loss: 1.877126305771526e-05
step: 130, loss: 9.602480713510886e-05
step: 140, loss: 0.00043559801997616887
step: 150, loss: 7.300647121155635e-05
step: 160, loss: 0.0007039180491119623
step: 170, loss: 6.187454710016027e-05
step: 180, loss: 0.00011168990022270009
step: 190, loss: 3.0446028176811524e-05
step: 200, loss: 0.0003781014820560813
step: 210, loss: 0.0004777007270604372
step: 220, loss: 2.4292034140671603e-05
step: 230, loss: 0.0006596475723199546
step: 240, loss: 5.539008998312056e-05
step: 250, loss: 3.132867641397752e-05
step: 260, loss: 1.6923704606597312e-05
step: 270, loss: 0.0007183292182162404
step: 280, loss: 2.013849916693289e-05
step: 290, loss: 0.001272786408662796
step: 300, loss: 3.053032560274005e-05
step: 310, loss: 3.126361480099149e-05
step: 320, loss: 0.0011786683462560177
step: 330, loss: 3.960356843890622e-05
step: 340, loss: 8.491990593029186e-05
step: 350, loss: 0.00010777533316286281
step: 360, loss: 0.00029157500830478966
step: 370, loss: 2.3222866730066016e-05
step: 380, loss: 0.0001956828637048602
step: 390, loss: 6.354604556690902e-05
step: 400, loss: 2.4601367840659805e-05
step: 410, loss: 1.9810711819445714e-05
step: 420, loss: 1.9948512999690138e-05
step: 430, loss: 5.2795501687796786e-05
step: 440, loss: 1.5336827345890924e-05
step: 450, loss: 4.448627441888675e-05
step: 460, loss: 3.075831409660168e-05
step: 470, loss: 0.0024600077886134386
step: 480, loss: 3.0228968171286397e-05
step: 490, loss: 7.661060953978449e-05
step: 500, loss: 0.02089032344520092
step: 510, loss: 2.657541881490033e-05
step: 520, loss: 3.668071803986095e-05
step: 530, loss: 1.9505325326463208e-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=0.9233698238933842, f1=0.9265750828991001,
```

```
best_f1=0.9285051067780873
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 2.1725267288275063e-05
step: 10, loss: 0.0003144262300338596
step: 20, loss: 3.824610757874325e-05
step: 30, loss: 2.4135553758242168e-05
step: 40, loss: 1.7650014342507347e-05
step: 50, loss: 1.5273488315870054e-05
step: 60, loss: 1.9587152564781718e-05
step: 70, loss: 7.142400863813236e-05
step: 80, loss: 3.5560129617806524e-05
step: 90, loss: 2.2093441657489166e-05
step: 100, loss: 2.4030297936405987e-05
step: 110, loss: 0.0001639237889321521
step: 120, loss: 0.0009357865783385932
step: 130, loss: 2.46299387072213e-05
step: 140, loss: 1.754583172441926e-05
step: 150, loss: 4.158100637141615e-05
step: 160, loss: 2.427331310173031e-05
step: 170, loss: 2.0667481294367462e-05
step: 180, loss: 1.8980012100655586e-05
step: 190, loss: 2.8914999347762205e-05
step: 200, loss: 2.0391791622387245e-05
step: 210, loss: 1.8294573237653822e-05
step: 220, loss: 1.5228817574097775e-05
step: 230, loss: 2.7163076083525084e-05
step: 240, loss: 2.529788253013976e-05
step: 250, loss: 0.00014692310651298612
step: 260, loss: 1.9143875761074014e-05
step: 270, loss: 0.00011827475827885792
step: 280, loss: 5.0376398576190695e-05
step: 290, loss: 0.0001351411483483389
step: 300, loss: 0.0003460885491222143
step: 310, loss: 3.434856625972316e-05
step: 320, loss: 2.527542346797418e-05
step: 330, loss: 3.251263842685148e-05
step: 340, loss: 7.515389734180644e-05
step: 350, loss: 2.571858567534946e-05
step: 360, loss: 6.238286732695997e-05
step: 370, loss: 1.7728332750266418e-05
step: 380, loss: 2.6083484044647776e-05
step: 390, loss: 1.5482115486520343e-05
step: 400, loss: 0.0008217744180001318
step: 410, loss: 2.9569755497504957e-05
step: 420, loss: 0.00010940213542198762
step: 430, loss: 2.5986784748965874e-05
```



```

step: 440, loss: 0.03992190584540367
step: 450, loss: 0.00011113841173937544
step: 460, loss: 1.9646349755930714e-05
step: 470, loss: 9.589862020220608e-05
step: 480, loss: 2.6425504984217696e-05
step: 490, loss: 0.0036747436970472336
step: 500, loss: 0.00021542525792028755
step: 510, loss: 2.5457406081841327e-05
step: 520, loss: 0.00026824965607374907
step: 530, loss: 2.8094964363845065e-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=0.9243295019157088, f1=0.9225621414913958,
best_f1=0.9285051067780873
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
5742it [00:17, 335.65it/s]
load_f1 = 0.9276879162702188
real_f1 = 0.9255014326647565

```

```
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:11, 368.13it/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        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8732722997665405
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=0.2772277227722772, f1=0.27184466019417475,
best_f1=0.27184466019417475
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.3690846264362335
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.39215686274509803, f1=0.3188405797101449,
best_f1=0.3188405797101449
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.3322870135307312
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.4186046511627907, f1=0.34285714285714286,
best_f1=0.34285714285714286
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.35872238874435425
/usr/local/lib/python3.7/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=0.4166666666666667, f1=0.34285714285714286,
best_f1=0.34285714285714286
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2889218032360077
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.4444444444444445, f1=0.4, best_f1=0.4
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.23333169519901276

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.4799999999999999, f1=0.4444444444444444,
best_f1=0.4444444444444444
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.25851762294769287
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5, f1=0.4, best_f1=0.4
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.352406769990921
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5454545454545454, f1=0.3448275862068965,
best_f1=0.3448275862068965
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.17039251327514648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6, f1=0.4444444444444445, best_f1=0.4444444444444445
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1680360585451126
/usr/local/lib/python3.7/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=0.6, f1=0.4137931034482759, best_f1=0.4444444444444445
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1479111611843109
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6666666666666666, f1=0.4, best_f1=0.4
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.21724960207939148
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7200000000000001, f1=0.45, best_f1=0.45
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.061444222927093506
/usr/local/lib/python3.7/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=0.7200000000000001, f1=0.3888888888888889, best_f1=0.45
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1531502902507782
/usr/local/lib/python3.7/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=0.7200000000000001, f1=0.4, best_f1=0.45
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```

    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.22424361109733582
/usr/local/lib/python3.7/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=0.7200000000000001, f1=0.4, best_f1=0.45
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, 135204.27it/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 = 0.64
real_f1 = 0.6153846153846153
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, 373.89it/s]

```

4.1.2 DBLP-ACM - Running the matcher

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

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

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

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

Defaults for this optimization level are:

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


```

Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale              : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.7970069646835327
step: 10, loss: 0.4608150124549866
step: 20, loss: 0.5831175446510315
step: 30, loss: 0.44694218039512634
step: 40, loss: 0.20978543162345886
step: 50, loss: 0.0926455706357956
step: 60, loss: 0.08948671817779541
step: 70, loss: 0.16914696991443634
step: 80, loss: 0.1486378014087677
step: 90, loss: 0.047054655849933624
step: 100, loss: 0.2523473799228668
step: 110, loss: 0.0960240364074707
step: 120, loss: 0.083842433989048
step: 130, loss: 0.009856713004410267
step: 140, loss: 0.14980579912662506
step: 150, loss: 0.09610266238451004
step: 160, loss: 0.16438782215118408
step: 170, loss: 0.00789871346205473
step: 180, loss: 0.013577114790678024
step: 190, loss: 0.028494859114289284
step: 200, loss: 0.02011859603226185
step: 210, loss: 0.0037276307120919228
step: 220, loss: 0.00272244936786592

```

```

step: 230, loss: 0.014272110536694527
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.975, f1=0.9665513264129182, best_f1=0.9665513264129182
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.007634921930730343
step: 10, loss: 0.00435772305354476
step: 20, loss: 0.008439389988780022
step: 30, loss: 0.00882037915289402
step: 40, loss: 0.02593560889363289
step: 50, loss: 0.04550020024180412
step: 60, loss: 0.006560991518199444
step: 70, loss: 0.023727942258119583
step: 80, loss: 0.00727373082190752
step: 90, loss: 0.006259301211684942
step: 100, loss: 0.05497630685567856
step: 110, loss: 0.0290799792855978
step: 120, loss: 0.12786608934402466
step: 130, loss: 0.024216726422309875
step: 140, loss: 0.005102784372866154
step: 150, loss: 0.00861409679055214
step: 160, loss: 0.021581800654530525
step: 170, loss: 0.1002553179860115
step: 180, loss: 0.03800716996192932
step: 190, loss: 0.016421223059296608
step: 200, loss: 0.011815875768661499
step: 210, loss: 0.028987402096390724
step: 220, loss: 0.017733242362737656
step: 230, loss: 0.026393819600343704
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9773242630385486, f1=0.9807037457434733,
best_f1=0.9807037457434733
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.004176101181656122
step: 10, loss: 0.022318754345178604
step: 20, loss: 0.032390423119068146
step: 30, loss: 0.007735471241176128
step: 40, loss: 0.013064723461866379
step: 50, loss: 0.005501435603946447
step: 60, loss: 0.022905796766281128
step: 70, loss: 0.0013277768157422543
step: 80, loss: 0.0543406680226326
step: 90, loss: 0.006115665193647146
step: 100, loss: 0.007451507728546858
step: 110, loss: 0.010236659087240696
step: 120, loss: 0.0021712053567171097
step: 130, loss: 0.002416068920865655
step: 140, loss: 0.0017352859722450376
step: 150, loss: 0.060421206057071686
step: 160, loss: 0.0022906861267983913
step: 170, loss: 0.004401764366775751
step: 180, loss: 0.0037391153164207935
step: 190, loss: 0.034966208040714264
step: 200, loss: 0.0010816409485414624
step: 210, loss: 0.05728446692228317
step: 220, loss: 0.006499935872852802
step: 230, loss: 0.018539344891905785
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9831271091113611, f1=0.9761092150170648,
best_f1=0.9761092150170648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0030169545207172632
step: 10, loss: 0.0017201892333105206
step: 20, loss: 0.0007229043985716999
step: 30, loss: 0.0007031565182842314
step: 40, loss: 0.0016259064432233572
step: 50, loss: 0.0030598740559071302
step: 60, loss: 0.00038723181933164597
step: 70, loss: 0.029597390443086624
step: 80, loss: 0.10165674984455109
step: 90, loss: 0.08099701255559921
step: 100, loss: 0.012286374345421791
step: 110, loss: 0.009322094731032848

```

```

step: 120, loss: 0.006804138422012329
step: 130, loss: 0.02340376190841198
step: 140, loss: 0.0012563127093017101
step: 150, loss: 0.035963643342256546
step: 160, loss: 0.002141506876796484
step: 170, loss: 0.00931647140532732
step: 180, loss: 0.06346062570810318
step: 190, loss: 0.007277487777173519
step: 200, loss: 0.0005676081636920571
step: 210, loss: 0.00041685643373057246
step: 220, loss: 0.0005514947115443647
step: 230, loss: 0.0005297554307617247
/usr/local/lib/python3.7/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=0.9820627802690582, f1=0.9798206278026906,
best_f1=0.9761092150170648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001368651632219553
step: 10, loss: 0.05394107103347778
step: 20, loss: 0.0027621390763670206
step: 30, loss: 0.0005638605216518044
step: 40, loss: 0.0034385058097541332
step: 50, loss: 0.0005641977186314762
step: 60, loss: 0.0010448620887473226
step: 70, loss: 0.0004484706441871822
step: 80, loss: 0.022876067087054253
step: 90, loss: 0.01619269698858261
step: 100, loss: 0.1481112688779831
step: 110, loss: 0.0855092704296112
step: 120, loss: 0.0313417874276638
step: 130, loss: 0.015696166083216667
step: 140, loss: 0.0002496563538443297
step: 150, loss: 0.010861304588615894
step: 160, loss: 0.0010259042028337717
step: 170, loss: 0.10352035611867905
step: 180, loss: 0.005104575306177139
step: 190, loss: 0.07053906470537186
step: 200, loss: 0.010171627625823021
step: 210, loss: 0.0011701392941176891
step: 220, loss: 0.0005504561122506857
step: 230, loss: 0.0010455211158841848
/usr/local/lib/python3.7/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=0.9809203142536477, f1=0.9865470852017937,

```

```

best_f1=0.9761092150170648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006178119219839573
step: 10, loss: 0.0004901717766188085
step: 20, loss: 0.000810410943813622
step: 30, loss: 0.005330891814082861
step: 40, loss: 0.001835472765378654
step: 50, loss: 0.011200562119483948
step: 60, loss: 0.017005832865834236
step: 70, loss: 0.001267320360057056
step: 80, loss: 0.0034628764260560274
step: 90, loss: 0.0031312855426222086
step: 100, loss: 0.00029737743898294866
step: 110, loss: 0.060178741812705994
step: 120, loss: 0.025443656370043755
step: 130, loss: 0.005120507441461086
step: 140, loss: 0.0024562578182667494
step: 150, loss: 0.0004492850857786834
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 160, loss: 0.014980699867010117
step: 170, loss: 0.004551695194095373
step: 180, loss: 0.00033138308208435774
step: 190, loss: 0.0011002563405781984
step: 200, loss: 0.03029402531683445
step: 210, loss: 0.001168454997241497
step: 220, loss: 0.00021408029715530574
step: 230, loss: 0.000309816183289513
/usr/local/lib/python3.7/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=0.9786276715410572, f1=0.9831271091113611,
best_f1=0.9761092150170648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.14754679799079895
step: 10, loss: 0.00021687809203285724
step: 20, loss: 0.00024788020527921617
step: 30, loss: 0.000692406203597784
step: 40, loss: 0.008039913140237331
step: 50, loss: 0.00032162806019186974
step: 60, loss: 0.004341126419603825
step: 70, loss: 0.0005668599042110145
step: 80, loss: 0.0012131404364481568
step: 90, loss: 0.0015389740001410246
step: 100, loss: 0.0015985388308763504

```

```

step: 110, loss: 0.033462852239608765
step: 120, loss: 0.0006949612870812416
step: 130, loss: 0.0006647810805588961
step: 140, loss: 0.00018656678730621934
step: 150, loss: 0.015780426561832428
step: 160, loss: 0.00031415195553563535
step: 170, loss: 0.0025299161206930876
step: 180, loss: 0.0775507315993309
step: 190, loss: 0.00069497583899647
step: 200, loss: 0.004168496001511812
step: 210, loss: 0.00017013213073369116
step: 220, loss: 0.010310176759958267
step: 230, loss: 0.034460414201021194
/usr/local/lib/python3.7/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=0.9819004524886877, f1=0.975, best_f1=0.9761092150170648
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.027091076597571373
step: 10, loss: 0.05609351396560669
step: 20, loss: 0.00044856470776721835
step: 30, loss: 0.0002656739379744977
step: 40, loss: 0.0004633860953617841
step: 50, loss: 0.0030071763321757317
step: 60, loss: 0.00013291194045450538
step: 70, loss: 0.0001871781423687935
step: 80, loss: 0.002011653268709779
step: 90, loss: 0.00023408219567500055
step: 100, loss: 0.00019307431648485363
step: 110, loss: 0.005133455153554678
step: 120, loss: 0.0005509534967131913
step: 130, loss: 0.035158317536115646
step: 140, loss: 0.00011072739289375022
step: 150, loss: 0.00010035291779786348
step: 160, loss: 0.00012167022214271128
step: 170, loss: 0.0006341378903016448
step: 180, loss: 0.008724268525838852
step: 190, loss: 0.00285841035656631
step: 200, loss: 0.00037496149889193475
step: 210, loss: 0.0002744849189184606
step: 220, loss: 0.0007566564599983394
step: 230, loss: 0.035325899720191956
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:

```

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=0.9875424688561721, f1=0.9830124575311437,

best_f1=0.9830124575311437

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

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

step: 0, loss: 0.0003862614103127271

step: 10, loss: 0.0002825079718604684

step: 20, loss: 0.005851321388036013

step: 30, loss: 0.00043578213080763817

step: 40, loss: 0.020129036158323288

step: 50, loss: 0.0001512527378508821

step: 60, loss: 0.000662155041936785

step: 70, loss: 0.03474348783493042

step: 80, loss: 0.05187709629535675

step: 90, loss: 0.00036167463986203074

step: 100, loss: 0.006830052938312292

step: 110, loss: 0.003909408114850521

step: 120, loss: 0.06097404658794403

step: 130, loss: 0.000158352151629515

step: 140, loss: 0.00016564037650823593

step: 150, loss: 0.00010935168393189088

step: 160, loss: 0.000335354678099975

step: 170, loss: 0.0001804138591978699

step: 180, loss: 0.0007957952911965549

step: 190, loss: 0.00020378483168315142

step: 200, loss: 0.00018474982061889023

step: 210, loss: 0.0012227222323417664

step: 220, loss: 0.035336967557668686

step: 230, loss: 0.01236607413738966

/usr/local/lib/python3.7/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=0.9817351598173515, f1=0.971297359357061,

best_f1=0.9830124575311437

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

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

step: 0, loss: 0.00028068709070794284

step: 10, loss: 9.346839942736551e-05

step: 20, loss: 0.0020044920966029167

step: 30, loss: 0.0007390339160338044

step: 40, loss: 0.0005484019056893885

step: 50, loss: 0.08602184802293777

step: 60, loss: 0.056713178753852844

step: 70, loss: 0.0009587962995283306

```

step: 80, loss: 0.006999892182648182
step: 90, loss: 0.0005403279792517424
step: 100, loss: 0.0008811490260995924
step: 110, loss: 0.005602120887488127
step: 120, loss: 0.015396269969642162
step: 130, loss: 0.0015890095382928848
step: 140, loss: 0.024700680747628212
step: 150, loss: 0.01632932387292385
step: 160, loss: 0.012391213327646255
step: 170, loss: 0.0008212118991650641
step: 180, loss: 0.0006007282645441592
step: 190, loss: 0.0006004627794027328
step: 200, loss: 0.00020074439817108214
step: 210, loss: 0.00022641252144239843
step: 220, loss: 0.003936911933124065
step: 230, loss: 0.0014984685694798827
/usr/local/lib/python3.7/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=0.9819413092550789, f1=0.9752808988764046,
best_f1=0.9830124575311437
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0001325213088421151
step: 10, loss: 0.00018206555978395045
step: 20, loss: 0.00012474549293983728
step: 30, loss: 0.00043350463965907693
step: 40, loss: 0.00031139457132667303
step: 50, loss: 0.0011071652406826615
step: 60, loss: 0.00014078088861424476
step: 70, loss: 0.0036757495254278183
step: 80, loss: 0.00019905943190678954
step: 90, loss: 0.0001356842985842377
step: 100, loss: 0.00016995210899040103
step: 110, loss: 0.00014264191850088537
step: 120, loss: 0.0006018938147462904
step: 130, loss: 7.269238267326728e-05
step: 140, loss: 8.216944843297824e-05
step: 150, loss: 6.878124986542389e-05
step: 160, loss: 0.00011455884668976068
step: 170, loss: 0.015713095664978027
step: 180, loss: 0.03445926308631897
step: 190, loss: 0.0008948036702349782
step: 200, loss: 0.00013213002239353955
step: 210, loss: 0.0001442813954781741
step: 220, loss: 8.222654287237674e-05
step: 230, loss: 0.026766661554574966

```



```

/usr/local/lib/python3.7/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=0.9842342342342343, f1=0.978675645342312,
best_f1=0.9830124575311437
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.020150423049926758
step: 10, loss: 0.009155831299722195
step: 20, loss: 0.00012327427975833416
step: 30, loss: 0.00030390918254852295
step: 40, loss: 0.0005061107221990824
step: 50, loss: 8.07152537163347e-05
step: 60, loss: 0.01800522394478321
step: 70, loss: 6.925928755663335e-05
step: 80, loss: 9.702984243631363e-05
step: 90, loss: 9.326819417765364e-05
step: 100, loss: 6.234113243408501e-05
step: 110, loss: 0.0001373870181851089
step: 120, loss: 0.00010419494356028736
step: 130, loss: 0.002527746371924877
step: 140, loss: 0.00014727703819517046
step: 150, loss: 5.336439789971337e-05
step: 160, loss: 0.02959243766963482
step: 170, loss: 0.00227541895583272
step: 180, loss: 0.12606866657733917
step: 190, loss: 0.0010273040970787406
step: 200, loss: 0.006209535989910364
step: 210, loss: 8.057457307586446e-05
step: 220, loss: 0.019181283190846443
step: 230, loss: 0.0001477014011470601
/usr/local/lib/python3.7/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=0.9841269841269841, f1=0.9783352337514253,
best_f1=0.9830124575311437
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0036369168665260077
step: 10, loss: 8.427606371697038e-05
step: 20, loss: 0.00012889136269222945
step: 30, loss: 0.024865297600626945
step: 40, loss: 0.00041411316487938166
step: 50, loss: 8.802572847343981e-05
step: 60, loss: 0.0001309439685428515
step: 70, loss: 0.0001450338022550568

```

```

step: 80, loss: 8.475124923279509e-05
step: 90, loss: 7.529481808887795e-05
step: 100, loss: 0.00011426040146034211
step: 110, loss: 0.00045855631469748914
step: 120, loss: 8.518256800016388e-05
step: 130, loss: 0.00014395306061487645
step: 140, loss: 0.0010460671037435532
step: 150, loss: 0.003917008638381958
step: 160, loss: 8.84867476997897e-05
step: 170, loss: 0.00013877387391403317
step: 180, loss: 0.001091850339435041
step: 190, loss: 6.515980930998921e-05
step: 200, loss: 7.744521280983463e-05
step: 210, loss: 0.00034767616307362914
step: 220, loss: 5.938670801697299e-05
step: 230, loss: 7.32413973310031e-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=0.9830890642615557, f1=0.9807909604519773,
best_f1=0.9830124575311437
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00017715075227897614
step: 10, loss: 5.688861710950732e-05
step: 20, loss: 0.014371929690241814
step: 30, loss: 0.00013635666982736439
step: 40, loss: 0.00011165982868988067
step: 50, loss: 0.00011444515985203907
step: 60, loss: 7.538941281381994e-05
step: 70, loss: 7.092181476764381e-05
step: 80, loss: 7.946347614051774e-05
step: 90, loss: 0.00033685730886645615
step: 100, loss: 0.01680661365389824
step: 110, loss: 0.00016020453767850995
step: 120, loss: 0.0001930679427459836
step: 130, loss: 0.0006549394456669688
step: 140, loss: 0.0012243585661053658
step: 150, loss: 0.0002179076400352642
step: 160, loss: 5.421262176241726e-05
step: 170, loss: 7.891003042459488e-05
step: 180, loss: 8.803629316389561e-05
step: 190, loss: 0.010333222337067127
step: 200, loss: 6.998330354690552e-05
step: 210, loss: 0.001985553652048111
step: 220, loss: 0.00013607126311399043
step: 230, loss: 0.00014905490388628095

```

```

/usr/local/lib/python3.7/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=0.9841628959276018, f1=0.9794988610478361,
best_f1=0.9830124575311437
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.836589015321806e-05
step: 10, loss: 0.0001565443235449493
step: 20, loss: 0.00013936497271060944
step: 30, loss: 0.0007604506681673229
step: 40, loss: 8.755258022574708e-05
step: 50, loss: 8.806013647699729e-05
step: 60, loss: 0.00011297044693492353
step: 70, loss: 0.0001690552307991311
step: 80, loss: 0.023930784314870834
step: 90, loss: 4.137439464102499e-05
step: 100, loss: 0.0001412491110386327
step: 110, loss: 0.0020931167528033257
step: 120, loss: 0.09782321006059647
step: 130, loss: 0.00013765343464910984
step: 140, loss: 0.00039330546860583127
step: 150, loss: 0.0006717095384374261
step: 160, loss: 6.824052252341062e-05
step: 170, loss: 0.00012266592239029706
step: 180, loss: 0.0008833690080791712
step: 190, loss: 0.015266377478837967
step: 200, loss: 0.00019353063544258475
step: 210, loss: 0.006994095630943775
step: 220, loss: 0.00022073336003813893
step: 230, loss: 0.0003549413522705436
/usr/local/lib/python3.7/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=0.9841986455981941, f1=0.9808342728297633,
best_f1=0.9830124575311437
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

```

Defaults for this optimization level are:

```

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

```

```

loss_scale          : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled             : True
opt_level           : 02
cast_model_type     : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32  : True
master_weights      : True
loss_scale          : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2473it [00:07, 315.21it/s]
load_f1 = 0.9841628959276018
real_f1 = 0.9819004524886877
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, 352.03it/s]

```

4.1.3 DBLP-GoogleScholar - Running the matcher

```

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

```

```

--da entry_swap \
--da attr_del

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

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

```

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

Defaults for this optimization level are:

```

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

```

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

After processing overrides, optimization options are:

```

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

```

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

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

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

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

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

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

step: 0, loss: 0.7955784201622009

step: 10, loss: 0.42031317949295044

step: 20, loss: 0.47387516498565674

step: 30, loss: 0.4071016311645508

step: 40, loss: 0.3763653039932251

step: 50, loss: 0.18457384407520294

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: 60, loss: 0.22025829553604126

step: 70, loss: 0.14245253801345825

step: 80, loss: 0.14126057922840118

step: 90, loss: 0.11297115683555603

step: 100, loss: 0.3376676142215729

step: 110, loss: 0.04998636618256569

step: 120, loss: 0.07856583595275879

step: 130, loss: 0.0322941392660141

step: 140, loss: 0.1902209371328354

step: 150, loss: 0.03894586116075516

step: 160, loss: 0.15043804049491882

step: 170, loss: 0.2272070050239563

step: 180, loss: 0.1090407744050026

step: 190, loss: 0.04419432953000069

step: 200, loss: 0.07803870737552643

step: 210, loss: 0.12904012203216553

step: 220, loss: 0.14787134528160095

step: 230, loss: 0.16949228942394257

step: 240, loss: 0.055286239832639694

step: 250, loss: 0.08084213733673096

step: 260, loss: 0.024514274671673775

step: 270, loss: 0.032043956220149994

step: 280, loss: 0.11205644905567169

step: 290, loss: 0.1192479357123375

step: 300, loss: 0.08603735268115997

step: 310, loss: 0.08091378211975098

step: 320, loss: 0.1157255619764328

step: 330, loss: 0.16877205669879913

step: 340, loss: 0.08003778755664825

step: 350, loss: 0.08219488710165024

step: 360, loss: 0.052046336233615875

step: 370, loss: 0.1731915920972824

step: 380, loss: 0.1779443919658661

step: 390, loss: 0.03178108483552933

```

step: 400, loss: 0.022047916427254677
step: 410, loss: 0.032000355422496796
step: 420, loss: 0.028908712789416313
step: 430, loss: 0.15696655213832855
step: 440, loss: 0.12807121872901917
step: 450, loss: 0.03637754172086716
step: 460, loss: 0.056102216243743896
step: 470, loss: 0.23950788378715515
step: 480, loss: 0.19335544109344482
step: 490, loss: 0.12035925686359406
step: 500, loss: 0.025373447686433792
step: 510, loss: 0.07998889684677124
step: 520, loss: 0.14179851114749908
step: 530, loss: 0.19191694259643555
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.921173235563703, f1=0.9195612431444242,
best_f1=0.9195612431444242
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05502839386463165
step: 10, loss: 0.11986624449491501
step: 20, loss: 0.26070868968963623
step: 30, loss: 0.07438334077596664
step: 40, loss: 0.011126000434160233
step: 50, loss: 0.2028300017118454
step: 60, loss: 0.19468899071216583
step: 70, loss: 0.02862529270350933
step: 80, loss: 0.07853478938341141
step: 90, loss: 0.06580574810504913
step: 100, loss: 0.2220560908317566
step: 110, loss: 0.027510231360793114
step: 120, loss: 0.07849051803350449
step: 130, loss: 0.051912449300289154
step: 140, loss: 0.044416140764951706
step: 150, loss: 0.060319043695926666
step: 160, loss: 0.013235088437795639
step: 170, loss: 0.11345351487398148
step: 180, loss: 0.0390450693666935
step: 190, loss: 0.0059533086605370045
step: 200, loss: 0.07224451750516891
step: 210, loss: 0.016309987753629684

```

```

step: 220, loss: 0.17087817192077637
step: 230, loss: 0.007403685711324215
step: 240, loss: 0.1317843794822693
step: 250, loss: 0.07604395598173141
step: 260, loss: 0.04159407317638397
step: 270, loss: 0.03235120326280594
step: 280, loss: 0.1021442785859108
step: 290, loss: 0.15806350111961365
step: 300, loss: 0.020449001342058182
step: 310, loss: 0.07686494290828705
step: 320, loss: 0.1869315654039383
step: 330, loss: 0.01475820317864418
step: 340, loss: 0.012589446268975735
step: 350, loss: 0.04178300127387047
step: 360, loss: 0.08492573350667953
step: 370, loss: 0.0037937595043331385
step: 380, loss: 0.06196555495262146
step: 390, loss: 0.019306520000100136
step: 400, loss: 0.06902280449867249
step: 410, loss: 0.005261669401079416
step: 420, loss: 0.10180068761110306
step: 430, loss: 0.023869352415204048
step: 440, loss: 0.02581501565873623
step: 450, loss: 0.026667717844247818
step: 460, loss: 0.22803249955177307
step: 470, loss: 0.046458352357149124
step: 480, loss: 0.22042043507099152
step: 490, loss: 0.04845248535275459
step: 500, loss: 0.04219481348991394
step: 510, loss: 0.08115393668413162
step: 520, loss: 0.028962967917323112
step: 530, loss: 0.19224070012569427
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9239280774550485, f1=0.9292279411764707,
best_f1=0.9292279411764707
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.01728028990328312
step: 10, loss: 0.027981305494904518
step: 20, loss: 0.21003606915473938
step: 30, loss: 0.28525835275650024

```


step: 40, loss: 0.026203390210866928
step: 50, loss: 0.011712522245943546
step: 60, loss: 0.05104929208755493
step: 70, loss: 0.03950488194823265
step: 80, loss: 0.006008176598697901
step: 90, loss: 0.04518922418355942
step: 100, loss: 0.03272516280412674
step: 110, loss: 0.0554957278072834
step: 120, loss: 0.01577458158135414
step: 130, loss: 0.026860976591706276
step: 140, loss: 0.019592536613345146
step: 150, loss: 0.1328520029783249
step: 160, loss: 0.015488148666918278
step: 170, loss: 0.00794245209544897
step: 180, loss: 0.050464823842048645
step: 190, loss: 0.01616077683866024
step: 200, loss: 0.0698479562997818
step: 210, loss: 0.011875294148921967
step: 220, loss: 0.0070500788278877735
step: 230, loss: 0.009802952408790588
step: 240, loss: 0.007695743814110756
step: 250, loss: 0.023807602003216743
step: 260, loss: 0.023186735808849335
step: 270, loss: 0.02126910164952278
step: 280, loss: 0.011265375651419163
step: 290, loss: 0.031544312834739685
step: 300, loss: 0.05674927681684494
step: 310, loss: 0.12766066193580627
step: 320, loss: 0.08679884672164917
step: 330, loss: 0.003108114469796419
step: 340, loss: 0.0023143859580159187
step: 350, loss: 0.02475331537425518
step: 360, loss: 0.02363390289247036
step: 370, loss: 0.00722217233851552
step: 380, loss: 0.0026076557114720345
step: 390, loss: 0.004830462858080864
step: 400, loss: 0.01267283782362938
step: 410, loss: 0.02065104804933071
step: 420, loss: 0.040299396961927414
step: 430, loss: 0.017572563141584396
step: 440, loss: 0.18817421793937683
step: 450, loss: 0.11855810880661011
step: 460, loss: 0.08756614476442337
step: 470, loss: 0.02478472702205181
step: 480, loss: 0.006368940696120262
step: 490, loss: 0.018265321850776672
step: 500, loss: 0.08103322982788086
step: 510, loss: 0.0022999958600848913

```
step: 520, loss: 0.034903135150671005
step: 530, loss: 0.03242693841457367
/usr/local/lib/python3.7/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=0.9217230199166282, f1=0.9211502782931354,
best_f1=0.9292279411764707
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0073677897453308105
step: 10, loss: 0.00963241420686245
step: 20, loss: 0.0488218292593956
step: 30, loss: 0.0020478027872741222
step: 40, loss: 0.00674266554415226
step: 50, loss: 0.023764334619045258
step: 60, loss: 0.0022918565664440393
step: 70, loss: 0.03443237766623497
step: 80, loss: 0.022072767838835716
step: 90, loss: 0.0064543490298092365
step: 100, loss: 0.017227545380592346
step: 110, loss: 0.003404119284823537
step: 120, loss: 0.0013736519031226635
step: 130, loss: 0.06662895530462265
step: 140, loss: 0.016044318675994873
step: 150, loss: 0.0015420992858707905
step: 160, loss: 0.004930693190544844
step: 170, loss: 0.002441688906401396
step: 180, loss: 0.006486753933131695
step: 190, loss: 0.012238659895956516
step: 200, loss: 0.0037320279516279697
step: 210, loss: 0.06877954304218292
step: 220, loss: 0.04897331818938255
step: 230, loss: 0.1424228847026825
step: 240, loss: 0.028111808001995087
step: 250, loss: 0.019117146730422974
step: 260, loss: 0.06942031532526016
step: 270, loss: 0.030993063002824783
step: 280, loss: 0.013837764970958233
step: 290, loss: 0.11748771369457245
step: 300, loss: 0.0019108196720480919
step: 310, loss: 0.006516746245324612
step: 320, loss: 0.033175989985466
step: 330, loss: 0.08974044770002365
step: 340, loss: 0.07795307040214539
step: 350, loss: 0.006833699997514486
step: 360, loss: 0.034741297364234924
step: 370, loss: 0.03044167347252369
```

```

step: 380, loss: 0.01174152735620737
step: 390, loss: 0.027786951512098312
step: 400, loss: 0.027360429987311363
step: 410, loss: 0.08605580776929855
step: 420, loss: 0.023602791130542755
step: 430, loss: 0.011762093752622604
step: 440, loss: 0.044891852885484695
step: 450, loss: 0.03326606750488281
step: 460, loss: 0.00400531617924571
step: 470, loss: 0.01001544389873743
step: 480, loss: 0.0319565087556839
step: 490, loss: 0.023177314549684525
step: 500, loss: 0.019635234028100967
step: 510, loss: 0.08660906553268433
step: 520, loss: 0.06602005660533905
step: 530, loss: 0.008170838467776775
/usr/local/lib/python3.7/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=0.9207828518173344, f1=0.9191449814126395,
best_f1=0.9292279411764707
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014474057592451572
step: 10, loss: 0.06499206274747849
step: 20, loss: 0.0068426476791501045
step: 30, loss: 0.021097593009471893
step: 40, loss: 0.010707590728998184
step: 50, loss: 0.052607402205467224
step: 60, loss: 0.0014276271685957909
step: 70, loss: 0.01374279148876667
step: 80, loss: 0.003329511499032378
step: 90, loss: 0.00045277844765223563
step: 100, loss: 0.0008303250069729984
step: 110, loss: 0.0021972714457660913
step: 120, loss: 0.007310112938284874
step: 130, loss: 0.005780432373285294
step: 140, loss: 0.006190498825162649
step: 150, loss: 0.024308325722813606
step: 160, loss: 0.18604668974876404
step: 170, loss: 0.021814832463860512
step: 180, loss: 0.0009490991360507905
step: 190, loss: 0.0004254519590176642
step: 200, loss: 0.02258799970149994
step: 210, loss: 0.009739444591104984
step: 220, loss: 0.00036157379508949816
step: 230, loss: 0.0007138747023418546

```

```

step: 240, loss: 0.010495553724467754
step: 250, loss: 0.0009346645674668252
step: 260, loss: 0.0009740131790749729
step: 270, loss: 0.0617670938372612
step: 280, loss: 0.14520394802093506
step: 290, loss: 0.001036722445860505
step: 300, loss: 0.03041006252169609
step: 310, loss: 0.005682738963514566
step: 320, loss: 0.0031288459431380033
step: 330, loss: 0.0015673754969611764
step: 340, loss: 0.007901504635810852
step: 350, loss: 0.012915517203509808
step: 360, loss: 0.02494894526898861
step: 370, loss: 0.02122042328119278
step: 380, loss: 0.0889027789235115
step: 390, loss: 0.0214960016310215
step: 400, loss: 0.014600872993469238
step: 410, loss: 0.0016869929386302829
step: 420, loss: 0.008307109586894512
step: 430, loss: 0.004874926060438156
step: 440, loss: 0.005748686380684376
step: 450, loss: 0.004037962760776281
step: 460, loss: 0.0017006477573886514
step: 470, loss: 0.008309662342071533
step: 480, loss: 0.0066774082370102406
step: 490, loss: 0.0061725592240691185
step: 500, loss: 0.08314976841211319
step: 510, loss: 0.001924286363646388
step: 520, loss: 0.0005978651461191475
step: 530, loss: 0.0042381067760288715
/usr/local/lib/python3.7/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=0.9199812821712682, f1=0.9139382600561271,
best_f1=0.9292279411764707
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.020960599184036255
step: 10, loss: 0.011843333952128887
step: 20, loss: 0.034809429198503494
step: 30, loss: 0.00947560090571642
step: 40, loss: 0.002546170726418495
step: 50, loss: 0.017827710136771202
step: 60, loss: 0.006011674180626869
step: 70, loss: 0.18498019874095917
step: 80, loss: 0.0031263085547834635
step: 90, loss: 0.0010859109461307526

```

```
step: 100, loss: 0.07431524991989136
step: 110, loss: 0.0017694642301648855
step: 120, loss: 0.0026294116396456957
step: 130, loss: 0.019942523911595345
step: 140, loss: 0.003525946056470275
step: 150, loss: 0.01081470213830471
step: 160, loss: 0.004005725495517254
step: 170, loss: 0.003221604973077774
step: 180, loss: 0.0003604036464821547
step: 190, loss: 0.002028217539191246
step: 200, loss: 0.0003644716052804142
step: 210, loss: 0.0012170784175395966
step: 220, loss: 0.01568264327943325
step: 230, loss: 0.00028425990603864193
step: 240, loss: 0.1600343883037567
step: 250, loss: 0.07606668770313263
step: 260, loss: 0.011613596230745316
step: 270, loss: 0.002523513277992606
step: 280, loss: 0.07642457634210587
step: 290, loss: 0.0007759396103210747
step: 300, loss: 0.0032850911375135183
step: 310, loss: 0.0005366993136703968
step: 320, loss: 0.03800831735134125
step: 330, loss: 0.004220433533191681
step: 340, loss: 0.01851845160126686
step: 350, loss: 0.1186835840344429
step: 360, loss: 0.011186596006155014
step: 370, loss: 0.01240246556699276
step: 380, loss: 0.15966255962848663
step: 390, loss: 0.07436279952526093
step: 400, loss: 0.0010616866638883948
step: 410, loss: 0.02504033036530018
step: 420, loss: 0.015226715244352818
step: 430, loss: 0.0011472072219476104
step: 440, loss: 0.00552870100364089
step: 450, loss: 0.005100758746266365
step: 460, loss: 0.08273370563983917
step: 470, loss: 0.03828330338001251
step: 480, loss: 0.09602079540491104
step: 490, loss: 0.0009407291654497385
step: 500, loss: 0.007396458648145199
step: 510, loss: 0.00025707276654429734
step: 520, loss: 0.013033481314778328
step: 530, loss: 0.010623401962220669
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
```

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=0.9248826291079812, f1=0.9125295508274233,

best_f1=0.9125295508274233

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

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

step: 0, loss: 0.06355401873588562

step: 10, loss: 0.023615481331944466

step: 20, loss: 0.09669651091098785

step: 30, loss: 0.07717828452587128

step: 40, loss: 0.0006057134014554322

step: 50, loss: 0.018272489309310913

step: 60, loss: 0.16346706449985504

step: 70, loss: 0.03621954098343849

step: 80, loss: 0.002653437200933695

step: 90, loss: 0.008488746359944344

step: 100, loss: 0.000524458009749651

step: 110, loss: 0.01329924538731575

step: 120, loss: 0.0075948783196508884

step: 130, loss: 0.0005442058318294585

step: 140, loss: 0.0007691603968851268

step: 150, loss: 0.0001255115057574585

step: 160, loss: 0.0019435675349086523

step: 170, loss: 0.002047915942966938

step: 180, loss: 0.10137099772691727

step: 190, loss: 0.0004303802561480552

step: 200, loss: 0.001396602252498269

step: 210, loss: 0.00313957710750401

step: 220, loss: 0.0005877215880900621

step: 230, loss: 0.014574999921023846

step: 240, loss: 0.011364143341779709

step: 250, loss: 0.003553955350071192

step: 260, loss: 0.002294345060363412

step: 270, loss: 0.00013780518202111125

step: 280, loss: 0.015565305016934872

step: 290, loss: 0.028759315609931946

step: 300, loss: 0.0005648091901093721

step: 310, loss: 0.0004223008581902832

step: 320, loss: 0.03259947896003723

step: 330, loss: 0.006862992886453867

step: 340, loss: 0.01887023076415062

step: 350, loss: 0.000949875742662698

step: 360, loss: 0.02425697073340416

step: 370, loss: 0.0011591851944103837

step: 380, loss: 0.0165119431912899

step: 390, loss: 0.0002005489805014804

```

step: 400, loss: 0.00017075380310416222
step: 410, loss: 0.0004599115054588765
step: 420, loss: 0.000727064092643559
step: 430, loss: 0.00029192291549406946
step: 440, loss: 0.0005237767472863197
step: 450, loss: 0.007895123213529587
step: 460, loss: 0.017108161002397537
step: 470, loss: 0.07696841657161713
step: 480, loss: 0.02466624043881893
step: 490, loss: 0.008245046250522137
step: 500, loss: 0.0004758111899718642
step: 510, loss: 0.00237973197363317
step: 520, loss: 0.004100348800420761
step: 530, loss: 0.00015027458721306175
/usr/local/lib/python3.7/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=0.9245719574271172, f1=0.922654462242563,
best_f1=0.9125295508274233
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008703775238245726
step: 10, loss: 0.012402443215250969
step: 20, loss: 0.0011494825594127178
step: 30, loss: 0.13161902129650116
step: 40, loss: 0.00019951746799051762
step: 50, loss: 0.010491786524653435
step: 60, loss: 0.0004320464504417032
step: 70, loss: 0.031745605170726776
step: 80, loss: 0.0003081459435634315
step: 90, loss: 0.0006231727893464267
step: 100, loss: 0.023436663672327995
step: 110, loss: 0.0029383907094597816
step: 120, loss: 0.0002599081490188837
step: 130, loss: 0.0019396465504541993
step: 140, loss: 0.00015235270257107913
step: 150, loss: 0.001511838287115097
step: 160, loss: 0.00022199572413228452
step: 170, loss: 0.008742781355977058
step: 180, loss: 0.0011618408607318997
step: 190, loss: 0.0016451150877401233
step: 200, loss: 0.0006633515004068613
step: 210, loss: 0.15577217936515808
step: 220, loss: 0.009451426565647125
step: 230, loss: 0.0005897362716495991
step: 240, loss: 0.0006409207708202302
step: 250, loss: 0.028111249208450317

```

```

step: 260, loss: 0.12628932297229767
step: 270, loss: 0.00016741144645493478
step: 280, loss: 0.005518116056919098
step: 290, loss: 0.011198014952242374
step: 300, loss: 0.001178354024887085
step: 310, loss: 0.014321534894406796
step: 320, loss: 0.005594328977167606
step: 330, loss: 0.004956028889864683
step: 340, loss: 0.00018234406888950616
step: 350, loss: 0.04782091826200485
step: 360, loss: 0.0008235264685936272
step: 370, loss: 0.0012319774832576513
step: 380, loss: 0.013598381541669369
step: 390, loss: 0.0002382636594120413
step: 400, loss: 0.040025338530540466
step: 410, loss: 0.00010616692452458665
step: 420, loss: 7.919460767880082e-05
step: 430, loss: 0.00012843107106164098
step: 440, loss: 0.0007699710549786687
step: 450, loss: 0.0011405454715713859
step: 460, loss: 0.0017073273193091154
step: 470, loss: 0.00019972513837274164
step: 480, loss: 0.00010579961963230744
step: 490, loss: 0.0022266923915594816
step: 500, loss: 0.0014024007832631469
step: 510, loss: 0.005092705134302378
step: 520, loss: 0.026563555002212524
step: 530, loss: 0.00011915442155441269
/usr/local/lib/python3.7/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=0.921803127874885, f1=0.9172727272727272,
best_f1=0.9125295508274233
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00037999110645614564
step: 10, loss: 0.00020458736980799586
step: 20, loss: 0.00020093581406399608
step: 30, loss: 0.00275394506752491
step: 40, loss: 0.024356072768568993
step: 50, loss: 0.0005385664408095181
step: 60, loss: 0.0008203371544368565
step: 70, loss: 0.07751806825399399
step: 80, loss: 0.0024272738955914974
step: 90, loss: 0.0031584748066961765
step: 100, loss: 0.0004507861449383199
step: 110, loss: 0.00216642115265131

```



```
step: 120, loss: 0.0003714541962835938
step: 130, loss: 0.0002456220390740782
step: 140, loss: 0.024898430332541466
step: 150, loss: 0.0001630897168070078
step: 160, loss: 0.00034299661638215184
step: 170, loss: 0.00742041552439332
step: 180, loss: 6.121354817878455e-05
step: 190, loss: 0.1292932778596878
step: 200, loss: 0.0010209075408056378
step: 210, loss: 0.00012533088738564402
step: 220, loss: 0.00824772659689188
step: 230, loss: 0.003516610711812973
step: 240, loss: 0.0016188285080716014
step: 250, loss: 0.0010679435217753053
step: 260, loss: 0.007653530221432447
step: 270, loss: 0.03308752551674843
step: 280, loss: 0.0002292136923642829
step: 290, loss: 0.000327714515151456
step: 300, loss: 0.0024537756107747555
step: 310, loss: 0.0004255037638358772
step: 320, loss: 5.652765685226768e-05
step: 330, loss: 0.0006864702445454895
step: 340, loss: 0.016491878777742386
step: 350, loss: 0.0008405677508562803
step: 360, loss: 0.03362152352929115
step: 370, loss: 0.012261278927326202
step: 380, loss: 0.0011649278458207846
step: 390, loss: 0.005035157781094313
step: 400, loss: 0.004887253977358341
step: 410, loss: 0.006562519818544388
step: 420, loss: 0.0003949754172936082
step: 430, loss: 8.561169670429081e-05
step: 440, loss: 0.00011001512757502496
step: 450, loss: 7.678294059587643e-05
step: 460, loss: 0.0008312455611303449
step: 470, loss: 6.866586772957817e-05
step: 480, loss: 0.01319645531475544
step: 490, loss: 8.241435716627166e-05
step: 500, loss: 0.013293418101966381
step: 510, loss: 0.0012918063439428806
step: 520, loss: 5.667206642101519e-05
step: 530, loss: 4.91785685881041e-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=0.9220657276995305, f1=0.9193245778611632,
best_f1=0.9125295508274233
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

An input tensor was not cuda.

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

```
step: 0, loss: 0.027579009532928467
step: 10, loss: 0.001526037696748972
step: 20, loss: 0.00013192383630666882
step: 30, loss: 0.0003047026402782649
step: 40, loss: 0.00042882972047664225
step: 50, loss: 0.00034518010215833783
step: 60, loss: 0.00035604709410108626
step: 70, loss: 0.00223427452147007
step: 80, loss: 0.004727678373456001
step: 90, loss: 0.0025026376824826
step: 100, loss: 0.006413437891751528
step: 110, loss: 0.059533171355724335
step: 120, loss: 0.0009779923129826784
step: 130, loss: 8.533152868039906e-05
step: 140, loss: 0.00018130496027879417
step: 150, loss: 0.00019452490960247815
step: 160, loss: 0.0003564954095054418
step: 170, loss: 0.0016411372926086187
step: 180, loss: 0.019438473507761955
step: 190, loss: 0.00020704248163383454
step: 200, loss: 0.002621862106025219
step: 210, loss: 0.004051329102367163
step: 220, loss: 0.000529819808434695
step: 230, loss: 0.0034209550358355045
step: 240, loss: 0.0005800186772830784
step: 250, loss: 0.00088261638302356
step: 260, loss: 0.001111663761548698
step: 270, loss: 4.4236101530259475e-05
step: 280, loss: 5.047550439485349e-05
step: 290, loss: 0.00011086911399615929
step: 300, loss: 0.02431604452431202
step: 310, loss: 0.00023289975069928914
step: 320, loss: 0.00020571405184455216
step: 330, loss: 0.010984192602336407
step: 340, loss: 0.0006766152800992131
step: 350, loss: 0.0006180944037623703
step: 360, loss: 0.00026455509942024946
step: 370, loss: 0.0006705656996928155
step: 380, loss: 0.0004988090950064361
step: 390, loss: 0.00029593674116767943
step: 400, loss: 0.0002132552908733487
step: 410, loss: 8.959665865404531e-05
step: 420, loss: 0.0020022469107061625
step: 430, loss: 6.157015013741329e-05
step: 440, loss: 0.0018555226270109415
step: 450, loss: 0.001094197854399681
```

```
step: 460, loss: 0.0003647622070275247
step: 470, loss: 7.329149957513437e-05
step: 480, loss: 0.00010703341104090214
step: 490, loss: 0.09218306094408035
step: 500, loss: 0.015034765005111694
step: 510, loss: 0.00139217684045434
step: 520, loss: 0.0022135782055556774
step: 530, loss: 0.00017759772890713066
/usr/local/lib/python3.7/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=0.9235100891600189, f1=0.9155261915998112,
best_f1=0.9125295508274233
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003687302232719958
step: 10, loss: 0.10752994567155838
step: 20, loss: 0.0010132647585123777
step: 30, loss: 0.0018882954027503729
step: 40, loss: 0.05327347666025162
step: 50, loss: 0.0003815604140982032
step: 60, loss: 0.0001707909832475707
step: 70, loss: 0.00010088561975862831
step: 80, loss: 9.864060120889917e-05
step: 90, loss: 0.00010961407679133117
step: 100, loss: 0.0003121269983239472
step: 110, loss: 0.005880335811525583
step: 120, loss: 0.0002740375348366797
step: 130, loss: 0.0006200013449415565
step: 140, loss: 8.316728053614497e-05
step: 150, loss: 0.0006417504628188908
step: 160, loss: 0.0964464619755745
step: 170, loss: 0.32731005549430847
step: 180, loss: 0.00019173885812051594
step: 190, loss: 0.02284524030983448
step: 200, loss: 0.0010317453416064382
step: 210, loss: 0.00011305115185678005
step: 220, loss: 0.00043598480988293886
step: 230, loss: 0.00020420821965672076
step: 240, loss: 0.0001391430851072073
step: 250, loss: 0.0006822270224802196
step: 260, loss: 0.0011735553853213787
step: 270, loss: 0.0018406064482405782
step: 280, loss: 0.0008741677738726139
step: 290, loss: 0.0001855457085184753
step: 300, loss: 0.04116227477788925
step: 310, loss: 0.019500385969877243
```

```

step: 320, loss: 0.0005247541703283787
step: 330, loss: 0.002695366507396102
step: 340, loss: 0.002074831398203969
step: 350, loss: 0.006361977197229862
step: 360, loss: 0.0005700482288375497
step: 370, loss: 0.0001957538042915985
step: 380, loss: 0.00024987838696688414
step: 390, loss: 0.0006770231411792338
step: 400, loss: 0.00011456550419097766
step: 410, loss: 0.008402147330343723
step: 420, loss: 0.0019053955329582095
step: 430, loss: 0.0007354284171015024
step: 440, loss: 0.0001408050738973543
step: 450, loss: 0.0002726529201027006
step: 460, loss: 0.015565915033221245
step: 470, loss: 0.005033029243350029
step: 480, loss: 0.010986214503645897
step: 490, loss: 0.14686399698257446
step: 500, loss: 0.0016016026493161917
step: 510, loss: 0.00012485698971431702
step: 520, loss: 0.00018791636102832854
step: 530, loss: 0.00014121909043751657
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9257037378864791, f1=0.9200367647058824,
best_f1=0.9200367647058824
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0010323720052838326
step: 10, loss: 0.0004188968741800636
step: 20, loss: 6.694950570818037e-05
step: 30, loss: 0.0010709023335948586
step: 40, loss: 0.00018126526265405118
step: 50, loss: 0.0003134059370495379
step: 60, loss: 6.273394683375955e-05
step: 70, loss: 0.00857962854206562
step: 80, loss: 0.0007025696686469018
step: 90, loss: 0.0002616701240185648
step: 100, loss: 0.0007057649199850857
step: 110, loss: 0.0017978294054046273
step: 120, loss: 0.00032346995430998504
step: 130, loss: 0.00027674841112457216

```

```

step: 140, loss: 5.552199581870809e-05
step: 150, loss: 0.006743029225617647
step: 160, loss: 0.0013259764527902007
step: 170, loss: 0.00016020724433474243
step: 180, loss: 0.0001843395730247721
step: 190, loss: 8.559686102671549e-05
step: 200, loss: 0.00022594176698476076
step: 210, loss: 0.0009363693534396589
step: 220, loss: 6.902397581143305e-05
step: 230, loss: 0.0012034939136356115
step: 240, loss: 0.00015865860041230917
step: 250, loss: 7.32325206627138e-05
step: 260, loss: 0.00040007373900152743
step: 270, loss: 0.006512646097689867
step: 280, loss: 4.097577402717434e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
step: 290, loss: 0.0016245123697444797
step: 300, loss: 0.00046735661453567445
step: 310, loss: 0.00014187903434503824
step: 320, loss: 3.4911387047031894e-05
step: 330, loss: 0.00022008341329637915
step: 340, loss: 0.0046402825973927975
step: 350, loss: 0.00044264327152632177
step: 360, loss: 0.0008590344223193824
step: 370, loss: 0.13328836858272552
step: 380, loss: 0.0003153994621243328
step: 390, loss: 4.784395423484966e-05
step: 400, loss: 6.548461533384398e-05
step: 410, loss: 0.003950129263103008
step: 420, loss: 0.011501109227538109
step: 430, loss: 0.0003519164747558534
step: 440, loss: 0.0038144357968121767
step: 450, loss: 0.00012642476940527558
step: 460, loss: 0.0007145822746679187
step: 470, loss: 0.00012119534949306399
step: 480, loss: 0.00048646339564584196
step: 490, loss: 0.0018342185067012906
step: 500, loss: 0.0052840313874185085
step: 510, loss: 0.00046620145440101624
step: 520, loss: 0.00015325444110203534
step: 530, loss: 0.00010399709572084248
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9260299625468165, f1=0.9141248240262787,  
best_f1=0.9141248240262787  
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:  
An input tensor was not cuda.  
    warnings.warn("An input tensor was not cuda.")  
step: 0, loss: 0.0006587684038095176  
step: 10, loss: 5.0224200094817206e-05  
step: 20, loss: 0.006092086900025606  
step: 30, loss: 0.024843961000442505  
step: 40, loss: 0.0006108388188295066  
step: 50, loss: 8.029817399801686e-05  
step: 60, loss: 0.0011436165077611804  
step: 70, loss: 9.242996020475402e-05  
step: 80, loss: 0.0002504721051082015  
step: 90, loss: 0.0005568453343585134  
step: 100, loss: 0.00012653434532694519  
step: 110, loss: 0.0002653022820595652  
step: 120, loss: 0.0005268729291856289  
step: 130, loss: 0.0009097443544305861  
step: 140, loss: 0.06775165349245071  
step: 150, loss: 0.0029704307671636343  
step: 160, loss: 0.00010782874596770853  
step: 170, loss: 0.012014658190310001  
step: 180, loss: 0.00030985960620455444  
step: 190, loss: 0.0007169469608925283  
step: 200, loss: 0.0025633773766458035  
step: 210, loss: 0.0029660318978130817  
step: 220, loss: 0.0002234095591120422  
step: 230, loss: 0.0001717594568617642  
step: 240, loss: 5.5285847338382155e-05  
step: 250, loss: 0.005250425543636084  
step: 260, loss: 0.0008235286804847419  
step: 270, loss: 0.00018971752433571965  
step: 280, loss: 0.00021815308718942106  
step: 290, loss: 0.00023434311151504517  
step: 300, loss: 0.01587671972811222  
step: 310, loss: 0.0006694573676213622  
step: 320, loss: 7.622158591402695e-05  
step: 330, loss: 0.00018192151037510484  
step: 340, loss: 5.7607969210948795e-05  
step: 350, loss: 0.01818062551319599  
step: 360, loss: 9.072662942344323e-05  
step: 370, loss: 0.0001805234351195395  
step: 380, loss: 0.0006163078360259533  
step: 390, loss: 0.00031567647238262  
step: 400, loss: 4.673846342484467e-05  
step: 410, loss: 0.00038481727824546397  
step: 420, loss: 0.0005304788937792182
```

```
step: 430, loss: 0.000191616898518987
step: 440, loss: 0.0009372466593049467
step: 450, loss: 0.0010087109403684735
step: 460, loss: 2.9507309591281228e-05
step: 470, loss: 0.0036507712211459875
step: 480, loss: 0.0012903898023068905
step: 490, loss: 0.0002504984149709344
step: 500, loss: 6.876633415231481e-05
step: 510, loss: 0.00029540975810959935
step: 520, loss: 0.00020077958470210433
step: 530, loss: 0.00015016157703939825
/usr/local/lib/python3.7/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=0.9253592953175707, f1=0.9178966789667896,
best_f1=0.9141248240262787
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0002881193649955094
step: 10, loss: 8.410611189901829e-05
step: 20, loss: 0.00026504049310460687
step: 30, loss: 0.0038213329389691353
step: 40, loss: 0.0015415472444146872
step: 50, loss: 0.004691218491643667
step: 60, loss: 0.00023044856789056212
step: 70, loss: 0.00011038577213184908
step: 80, loss: 0.00019638193771243095
step: 90, loss: 9.023195161717013e-05
step: 100, loss: 0.007168244104832411
step: 110, loss: 0.00021927083435002714
step: 120, loss: 6.747312727384269e-05
step: 130, loss: 0.00024946077610366046
step: 140, loss: 0.004128933884203434
step: 150, loss: 0.030593400821089745
step: 160, loss: 0.002049729460850358
step: 170, loss: 0.013610472902655602
step: 180, loss: 0.0015339351957663894
step: 190, loss: 0.0028779059648513794
step: 200, loss: 0.06330292671918869
step: 210, loss: 0.0018719668732956052
step: 220, loss: 0.0005345907411538064
step: 230, loss: 0.00016583978140261024
step: 240, loss: 0.0028112924192100763
step: 250, loss: 7.391845429083332e-05
step: 260, loss: 6.972365372348577e-05
step: 270, loss: 0.005126702133566141
step: 280, loss: 0.0011058463715016842
```

```

step: 290, loss: 0.0052427309565246105
step: 300, loss: 0.0010556394699960947
step: 310, loss: 0.006263194605708122
step: 320, loss: 0.0011070732725784183
step: 330, loss: 0.0001344298798358068
step: 340, loss: 0.0003091772668994963
step: 350, loss: 0.0043135168962180614
step: 360, loss: 0.027088278904557228
step: 370, loss: 0.009599973447620869
step: 380, loss: 3.0166376745910384e-05
step: 390, loss: 9.810402116272599e-05
step: 400, loss: 4.586783688864671e-05
step: 410, loss: 0.00012408618931658566
step: 420, loss: 0.0030449600890278816
step: 430, loss: 0.0011152031365782022
step: 440, loss: 3.417465268285014e-05
step: 450, loss: 8.137078111758456e-05
step: 460, loss: 0.00030210489057935774
step: 470, loss: 4.499805436353199e-05
step: 480, loss: 6.825351738370955e-05
step: 490, loss: 9.018884156830609e-05
step: 500, loss: 0.0008322332287207246
step: 510, loss: 0.0014261876931414008
step: 520, loss: 0.0010695880046114326
step: 530, loss: 2.9639692002092488e-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=0.9253871421867668, f1=0.9147141518275539,
best_f1=0.9141248240262787
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.810790702933446e-05
step: 10, loss: 5.671284088748507e-05
step: 20, loss: 0.0010644919238984585
step: 30, loss: 0.00022086125682108104
step: 40, loss: 6.359074177453294e-05
step: 50, loss: 0.0016428359085693955
step: 60, loss: 0.00018064974574372172
step: 70, loss: 5.0510894652688876e-05
step: 80, loss: 0.00021881872089579701
step: 90, loss: 2.5640265448600985e-05
step: 100, loss: 6.895157275721431e-05
step: 110, loss: 0.009987089782953262
step: 120, loss: 0.0003437118139117956
step: 130, loss: 0.006746875122189522
step: 140, loss: 2.6828745831153356e-05

```



```

step: 150, loss: 0.0013068699045106769
step: 160, loss: 4.6290995669551194e-05
step: 170, loss: 0.0009486345807090402
step: 180, loss: 0.0014505151193588972
step: 190, loss: 0.0009381171548739076
step: 200, loss: 0.00017806874529924244
step: 210, loss: 0.00020391843281686306
step: 220, loss: 0.0002737520553637296
step: 230, loss: 0.0002670272660907358
step: 240, loss: 4.6097055019345134e-05
step: 250, loss: 0.00018750455637928098
step: 260, loss: 0.0003597138274926692
step: 270, loss: 0.00010396588913863525
step: 280, loss: 0.00016577185306232423
step: 290, loss: 0.0029061392415314913
step: 300, loss: 0.0010602576658129692
step: 310, loss: 0.00015434603847097605
step: 320, loss: 0.0004771971725858748
step: 330, loss: 0.0022869908716529608
step: 340, loss: 5.420598972705193e-05
step: 350, loss: 0.000889733899384737
step: 360, loss: 6.502051110146567e-05
step: 370, loss: 0.003945333417505026
step: 380, loss: 2.6016810807050206e-05
step: 390, loss: 0.00041275532566942275
step: 400, loss: 0.00028938031755387783
step: 410, loss: 0.00015739895752631128
step: 420, loss: 3.938521695090458e-05
step: 430, loss: 6.19150887359865e-05
step: 440, loss: 0.008073696866631508
step: 450, loss: 0.004011419136077166
step: 460, loss: 4.1184342990163714e-05
step: 470, loss: 0.00011393966997275129
step: 480, loss: 0.011002080515027046
step: 490, loss: 0.00021762892720289528
step: 500, loss: 0.0002997090632561594
step: 510, loss: 0.00737281097099185
step: 520, loss: 0.012196332216262817
step: 530, loss: 2.5413361072423868e-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=0.9258920402561757, f1=0.9170068027210884,
best_f1=0.9141248240262787
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

```

```

Defaults for this optimization level are:
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled          : True
opt_level        : 02
cast_model_type  : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32 : True
master_weights   : True
loss_scale       : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
5742it [00:15, 359.71it/s]
load_f1 = 0.9255269320843091
real_f1 = 0.9237961664329126
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, 357.31it/s]

```

4.1.4 Amazon-Google - Running the matcher

```

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

```

```

--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \
--dk product \
--summarize \
--da entry_swap \
--da attr_del

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

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

```

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

Defaults for this optimization level are:

```

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

```

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

After processing overrides, optimization options are:

```

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

```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8407125473022461
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.06752308458089828
step: 20, loss: 0.3839276134967804
step: 30, loss: 0.37384578585624695
step: 40, loss: 0.516684353351593
step: 50, loss: 0.3022063672542572
step: 60, loss: 0.36567309498786926
step: 70, loss: 0.2542398273944855
step: 80, loss: 0.27778172492980957
step: 90, loss: 0.4233831465244293
step: 100, loss: 0.19731467962265015
step: 110, loss: 0.28298431634902954
step: 120, loss: 0.205354705452919
step: 130, loss: 0.20931588113307953
step: 140, loss: 0.28658580780029297
step: 150, loss: 0.22721755504608154
step: 160, loss: 0.2153189331293106
step: 170, loss: 0.11541967839002609
step: 180, loss: 0.22031141817569733
step: 190, loss: 0.24251779913902283
step: 200, loss: 0.154466450214386
step: 210, loss: 0.39955660700798035
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5377532228360958, f1=0.5658914728682171,
best_f1=0.5658914728682171
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

step: 0, loss: 0.063423290848732
step: 10, loss: 0.08114267140626907
step: 20, loss: 0.22895555198192596
step: 30, loss: 0.13732820749282837
step: 40, loss: 0.10063446313142776
step: 50, loss: 0.21042153239250183
step: 60, loss: 0.03206183388829231
step: 70, loss: 0.21668455004692078
step: 80, loss: 0.21115891635417938
step: 90, loss: 0.12763303518295288
step: 100, loss: 0.0717698261141777
step: 110, loss: 0.07335543632507324
step: 120, loss: 0.24166108667850494
step: 130, loss: 0.22577182948589325
step: 140, loss: 0.2815975248813629
step: 150, loss: 0.1401962786912918
step: 160, loss: 0.06747660040855408
step: 170, loss: 0.15254521369934082
step: 180, loss: 0.2797551155090332
step: 190, loss: 0.23293060064315796
step: 200, loss: 0.23557604849338531
step: 210, loss: 0.28393682837486267
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5845511482254697, f1=0.6063157894736841,
best_f1=0.6063157894736841
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.166345477104187
step: 10, loss: 0.2561899721622467
step: 20, loss: 0.3744676113128662
step: 30, loss: 0.06538961827754974
step: 40, loss: 0.10981206595897675
step: 50, loss: 0.08378403633832932
step: 60, loss: 0.2890905737876892
step: 70, loss: 0.09622973203659058
step: 80, loss: 0.08105472475290298
step: 90, loss: 0.12100494652986526
step: 100, loss: 0.032574016600847244
step: 110, loss: 0.09317704290151596
step: 120, loss: 0.13043953478336334
step: 130, loss: 0.10575375705957413

```

```

step: 140, loss: 0.2897687256336212
step: 150, loss: 0.11830878257751465
step: 160, loss: 0.08903703093528748
step: 170, loss: 0.2918892502784729
step: 180, loss: 0.07268211245536804
step: 190, loss: 0.15061311423778534
step: 200, loss: 0.10307365655899048
step: 210, loss: 0.08858278393745422
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5975820379965459, f1=0.6140350877192983,
best_f1=0.6140350877192983
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.06019299477338791
step: 10, loss: 0.01841869205236435
step: 20, loss: 0.05940636992454529
step: 30, loss: 0.09011074900627136
step: 40, loss: 0.09428110718727112
step: 50, loss: 0.06881032139062881
step: 60, loss: 0.03633204102516174
step: 70, loss: 0.34071266651153564
step: 80, loss: 0.21875151991844177
step: 90, loss: 0.01671355403959751
step: 100, loss: 0.22570393979549408
step: 110, loss: 0.1860736757516861
step: 120, loss: 0.06496788561344147
step: 130, loss: 0.0721469596028328
step: 140, loss: 0.14287687838077545
step: 150, loss: 0.10731861740350723
step: 160, loss: 0.04141758754849434
step: 170, loss: 0.1030188649892807
step: 180, loss: 0.10657411068677902
step: 190, loss: 0.18275730311870575
step: 200, loss: 0.06981848180294037
step: 210, loss: 0.03647959232330322
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6145648312611013, f1=0.6010362694300517,
best_f1=0.6010362694300517
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.09943364560604095
step: 10, loss: 0.08599307388067245
step: 20, loss: 0.061226073652505875
step: 30, loss: 0.028658946976065636
step: 40, loss: 0.33676934242248535
step: 50, loss: 0.09327062219381332
step: 60, loss: 0.03615323454141617
step: 70, loss: 0.11090533435344696
step: 80, loss: 0.03372525796294212
step: 90, loss: 0.1370316743850708
step: 100, loss: 0.10261047631502151
step: 110, loss: 0.003295040689408779
step: 120, loss: 0.039806824177503586
step: 130, loss: 0.08696423470973969
step: 140, loss: 0.05302627384662628
step: 150, loss: 0.029378220438957214
step: 160, loss: 0.14459629356861115
step: 170, loss: 0.0809820219874382
step: 180, loss: 0.1328401267528534
step: 190, loss: 0.3346913754940033
step: 200, loss: 0.018381323665380478
step: 210, loss: 0.032462649047374725
/usr/local/lib/python3.7/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=0.6083499005964215, f1=0.6062992125984252,
best_f1=0.6010362694300517
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.03302633389830589
step: 10, loss: 0.06556283682584763
step: 20, loss: 0.036121342331171036
step: 30, loss: 0.13059931993484497
step: 40, loss: 0.013402897864580154
step: 50, loss: 0.010008086450397968
step: 60, loss: 0.02191259153187275
step: 70, loss: 0.03446757420897484
step: 80, loss: 0.13416559994220734
step: 90, loss: 0.11330954730510712
step: 100, loss: 0.017573686316609383
step: 110, loss: 0.030985087156295776

```

```

step: 120, loss: 0.03700101375579834
step: 130, loss: 0.049574051052331924
step: 140, loss: 0.0817955881357193
step: 150, loss: 0.030113443732261658
step: 160, loss: 0.12881650030612946
step: 170, loss: 0.045169148594141006
step: 180, loss: 0.024553529918193817
step: 190, loss: 0.12206046283245087
step: 200, loss: 0.01784234493970871
step: 210, loss: 0.03466685861349106
/usr/local/lib/python3.7/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=0.6000000000000001, f1=0.5940594059405941,
best_f1=0.6010362694300517
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04686685651540756
step: 10, loss: 0.016130708158016205
step: 20, loss: 0.07013621926307678
step: 30, loss: 0.003155176527798176
step: 40, loss: 0.1584596484899521
step: 50, loss: 0.12271956354379654
step: 60, loss: 0.18396714329719543
step: 70, loss: 0.008725129067897797
step: 80, loss: 0.09525804966688156
step: 90, loss: 0.025357017293572426
step: 100, loss: 0.05075313523411751
step: 110, loss: 0.16209077835083008
step: 120, loss: 0.16892974078655243
step: 130, loss: 0.018570920452475548
step: 140, loss: 0.008917650207877159
step: 150, loss: 0.020380735397338867
step: 160, loss: 0.039314739406108856
step: 170, loss: 0.05152701959013939
step: 180, loss: 0.0014692547265440226
step: 190, loss: 0.019106581807136536
step: 200, loss: 0.0689438134431839
step: 210, loss: 0.10864779353141785
/usr/local/lib/python3.7/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=0.5945945945945946, f1=0.6085192697768762,
best_f1=0.6010362694300517
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```



```

step: 0, loss: 0.05024644732475281
step: 10, loss: 0.033240050077438354
step: 20, loss: 0.11189496517181396
step: 30, loss: 0.010445772670209408
step: 40, loss: 0.10546117275953293
step: 50, loss: 0.0061635239981114864
step: 60, loss: 0.24758721888065338
step: 70, loss: 0.0030788914300501347
step: 80, loss: 0.009167301468551159
step: 90, loss: 0.23182345926761627
step: 100, loss: 0.002813430270180106
step: 110, loss: 0.022152898833155632
step: 120, loss: 0.015546022914350033
step: 130, loss: 0.03794444352388382
step: 140, loss: 0.01608104631304741
step: 150, loss: 0.04931309446692467
step: 160, loss: 0.14914770424365997
step: 170, loss: 0.09812337160110474
step: 180, loss: 0.029585042968392372
step: 190, loss: 0.02067737653851509
step: 200, loss: 0.08624966442584991
step: 210, loss: 0.1222933754324913
/usr/local/lib/python3.7/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=0.5815899581589957, f1=0.6107784431137724,
best_f1=0.6010362694300517
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008521256037056446
step: 10, loss: 0.026490245014429092
step: 20, loss: 0.03786792978644371
step: 30, loss: 0.5397061109542847
step: 40, loss: 0.014673377387225628
step: 50, loss: 0.029369128867983818
step: 60, loss: 0.009232109412550926
step: 70, loss: 0.06466840207576752
step: 80, loss: 0.06335756182670593
step: 90, loss: 0.0638638287782669
step: 100, loss: 0.02529752068221569
step: 110, loss: 0.041165392845869064
step: 120, loss: 0.023070547729730606
step: 130, loss: 0.010667084716260433
step: 140, loss: 0.03274824842810631
step: 150, loss: 0.022797400131821632
step: 160, loss: 0.010462412610650063
step: 170, loss: 0.009388132952153683

```

```

step: 180, loss: 0.025852875784039497
step: 190, loss: 0.04703165590763092
step: 200, loss: 0.08167719095945358
step: 210, loss: 0.06735630333423615
/usr/local/lib/python3.7/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=0.5792563600782779, f1=0.6007604562737644,
best_f1=0.6010362694300517
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005728791933506727
step: 10, loss: 0.001481713610701263
step: 20, loss: 0.02517053484916687
step: 30, loss: 0.01751568540930748
step: 40, loss: 0.060159020125865936
step: 50, loss: 0.0056884740479290485
step: 60, loss: 0.0024937489069998264
step: 70, loss: 0.049968019127845764
step: 80, loss: 0.0004539120418485254
step: 90, loss: 0.05213773623108864
step: 100, loss: 0.0025665322318673134
step: 110, loss: 0.00881543941795826
step: 120, loss: 0.0023604994639754295
step: 130, loss: 0.002349420450627804
step: 140, loss: 0.00033963628811761737
step: 150, loss: 0.04941387102007866
step: 160, loss: 0.008701527491211891
step: 170, loss: 0.082905113697052
step: 180, loss: 0.008357142098248005
step: 190, loss: 0.041704557836055756
step: 200, loss: 0.02348567359149456
step: 210, loss: 0.014754234813153744
/usr/local/lib/python3.7/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=0.6036217303822937, f1=0.5904761904761905,
best_f1=0.6010362694300517
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003970396239310503
step: 10, loss: 0.10921037942171097
step: 20, loss: 0.017672745510935783
step: 30, loss: 0.06923472136259079
step: 40, loss: 0.044467952102422714
step: 50, loss: 0.04201784357428551

```

```

step: 60, loss: 0.005866301245987415
step: 70, loss: 0.005059314426034689
step: 80, loss: 0.15412339568138123
step: 90, loss: 0.04606075584888458
step: 100, loss: 0.015212374739348888
step: 110, loss: 0.09127506613731384
step: 120, loss: 0.04352549463510513
step: 130, loss: 0.08420111984014511
step: 140, loss: 0.07522309571504593
step: 150, loss: 0.15095849335193634
step: 160, loss: 0.015735909342765808
step: 170, loss: 0.06958548724651337
step: 180, loss: 0.013164990581572056
step: 190, loss: 0.018180381506681442
step: 200, loss: 0.0013668496394529939
step: 210, loss: 0.0044548483565449715
/usr/local/lib/python3.7/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=0.5776805251641137, f1=0.6045548654244307,
best_f1=0.6010362694300517
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006472317036241293
step: 10, loss: 0.004981566686183214
step: 20, loss: 0.012062987312674522
step: 30, loss: 0.0017930142348632216
step: 40, loss: 0.15478092432022095
step: 50, loss: 0.036127083003520966
step: 60, loss: 0.0025262199342250824
step: 70, loss: 0.005250074900686741
step: 80, loss: 0.062031082808971405
step: 90, loss: 0.002093398245051503
step: 100, loss: 0.0010267465841025114
step: 110, loss: 0.04219867289066315
step: 120, loss: 0.007188720162957907
step: 130, loss: 0.01478843204677105
step: 140, loss: 0.007212395779788494
step: 150, loss: 0.0039623505435884
step: 160, loss: 0.025512082502245903
step: 170, loss: 0.004034234676510096
step: 180, loss: 0.006114352960139513
step: 190, loss: 0.21033547818660736
step: 200, loss: 0.02740662917494774
step: 210, loss: 0.005055742338299751
/usr/local/lib/python3.7/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=0.5762081784386617, f1=0.5960502692998205,
best_f1=0.6010362694300517
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0014486219733953476
step: 10, loss: 0.09015414118766785
step: 20, loss: 0.0015816211234778166
step: 30, loss: 0.050381138920784
step: 40, loss: 0.0018427257891744375
step: 50, loss: 0.002816881285980344
step: 60, loss: 0.0034572973381727934
step: 70, loss: 0.15392540395259857
step: 80, loss: 0.010639342479407787
step: 90, loss: 0.11980045586824417
step: 100, loss: 0.0532708615064621
step: 110, loss: 0.0024436581879854202
step: 120, loss: 0.0015520555898547173
step: 130, loss: 0.0016483139479532838
step: 140, loss: 0.01000103261321783
step: 150, loss: 0.00044963156688027084
step: 160, loss: 0.004627835471183062
step: 170, loss: 0.11329066753387451
step: 180, loss: 0.0401216484606266
step: 190, loss: 0.004415635485202074
step: 200, loss: 0.002663140185177326
step: 210, loss: 0.032154787331819534
/usr/local/lib/python3.7/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=0.5726495726495726, f1=0.5979797979797981,
best_f1=0.6010362694300517
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0017816380131989717
step: 10, loss: 0.006239140406250954
step: 20, loss: 0.001478813705034554
step: 30, loss: 0.005912215914577246
step: 40, loss: 0.0017320697661489248
step: 50, loss: 0.021564144641160965
step: 60, loss: 0.009012719616293907
step: 70, loss: 0.0396336130797863
step: 80, loss: 0.07792549580335617
step: 90, loss: 0.036682937294244766
step: 100, loss: 0.00435391440987587
step: 110, loss: 0.036581117659807205

```

```

step: 120, loss: 0.007735705003142357
step: 130, loss: 0.037472017109394073
step: 140, loss: 0.0008806964615359902
step: 150, loss: 0.022518163546919823
step: 160, loss: 0.0027189983520656824
step: 170, loss: 0.031433723866939545
step: 180, loss: 0.017352331429719925
step: 190, loss: 0.0031892594415694475
step: 200, loss: 0.0005276625161059201
step: 210, loss: 0.08063817024230957
/usr/local/lib/python3.7/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=0.5720620842572062, f1=0.5889830508474577,
best_f1=0.6010362694300517
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05137183889746666
step: 10, loss: 0.0040326728485524654
step: 20, loss: 0.0010598718654364347
step: 30, loss: 0.00491280946880579
step: 40, loss: 0.0020298901945352554
step: 50, loss: 0.00330036086961627
step: 60, loss: 0.004479255061596632
step: 70, loss: 0.004622230771929026
step: 80, loss: 0.06787492334842682
step: 90, loss: 0.09695304930210114
step: 100, loss: 0.019790656864643097
step: 110, loss: 0.0010157382348552346
step: 120, loss: 0.005446533672511578
step: 130, loss: 0.17221251130104065
step: 140, loss: 0.01122989784926176
step: 150, loss: 0.006716274190694094
step: 160, loss: 0.0036500822752714157
step: 170, loss: 0.028776170685887337
step: 180, loss: 0.0022271659690886736
step: 190, loss: 0.005928918719291687
step: 200, loss: 0.004684041254222393
step: 210, loss: 0.017878064885735512
/usr/local/lib/python3.7/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=0.5720720720720721, f1=0.5856832971800433,
best_f1=0.6010362694300517
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

```

```

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2293it [00:03, 577.00it/s]
load_f1 = 0.6176470588235294
real_f1 = 0.6165137614678899
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, 379.33it/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 02: FP16 training with FP32 batchnorm and FP32 master weights.

Defaults for this optimization level are:

```

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

```

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

After processing overrides, optimization options are:

```

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

```

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

```

Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8586640954017639
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.1688244640827179
step: 20, loss: 0.14965146780014038
step: 30, loss: 0.5023961663246155
step: 40, loss: 0.25870612263679504
step: 50, loss: 0.30391618609428406
step: 60, loss: 0.36467182636260986
step: 70, loss: 0.17248602211475372
step: 80, loss: 0.5151351690292358
step: 90, loss: 0.2545313537120819
step: 100, loss: 0.22389307618141174
step: 110, loss: 0.23551662266254425
step: 120, loss: 0.41765135526657104
step: 130, loss: 0.3409111499786377
step: 140, loss: 0.30698713660240173
step: 150, loss: 0.23161399364471436
step: 160, loss: 0.21489249169826508
step: 170, loss: 0.36906784772872925
step: 180, loss: 0.248758926987648
step: 190, loss: 0.1296476274728775
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.5984251968503937, f1=0.622107969151671,
best_f1=0.622107969151671
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.2450912743806839

```



```

step: 10, loss: 0.035985756665468216
step: 20, loss: 0.044938839972019196
step: 30, loss: 0.10438455641269684
step: 40, loss: 0.49674850702285767
step: 50, loss: 0.29629167914390564
step: 60, loss: 0.16251222789287567
step: 70, loss: 0.14960919320583344
step: 80, loss: 0.13029631972312927
step: 90, loss: 0.14828665554523468
step: 100, loss: 0.2754117250442505
step: 110, loss: 0.06514383852481842
step: 120, loss: 0.14869444072246552
step: 130, loss: 0.21280473470687866
step: 140, loss: 0.29829373955726624
step: 150, loss: 0.11362249404191971
step: 160, loss: 0.09192105382680893
step: 170, loss: 0.20828776061534882
step: 180, loss: 0.1092509999871254
step: 190, loss: 0.13889557123184204
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7204301075268816, f1=0.7724867724867726,
best_f1=0.7724867724867726
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.14783824980258942
step: 10, loss: 0.15942078828811646
step: 20, loss: 0.026600200682878494
step: 30, loss: 0.02265060320496559
step: 40, loss: 0.06380421668291092
step: 50, loss: 0.14682090282440186
step: 60, loss: 0.0707826092839241
step: 70, loss: 0.13550451397895813
step: 80, loss: 0.09561143815517426
step: 90, loss: 0.06534644216299057
step: 100, loss: 0.18834829330444336
step: 110, loss: 0.187440425157547
step: 120, loss: 0.04032707214355469
step: 130, loss: 0.08335226029157639
step: 140, loss: 0.0889582708477974
step: 150, loss: 0.01693115197122097
step: 160, loss: 0.06496761739253998

```

```

step: 170, loss: 0.02741851843893528
step: 180, loss: 0.09138471633195877
step: 190, loss: 0.1994338184595108
/usr/local/lib/python3.7/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=0.7125748502994012, f1=0.7289156626506026,
best_f1=0.7724867724867726
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.028811875730752945
step: 10, loss: 0.06400299817323685
step: 20, loss: 0.07387776672840118
step: 30, loss: 0.027776693925261497
step: 40, loss: 0.0024732942692935467
step: 50, loss: 0.01627759449183941
step: 60, loss: 0.13078920543193817
step: 70, loss: 0.0068000974133610725
step: 80, loss: 0.11299320310354233
step: 90, loss: 0.011976854875683784
step: 100, loss: 0.014980328269302845
step: 110, loss: 0.08504234254360199
step: 120, loss: 0.11845586448907852
step: 130, loss: 0.19996197521686554
step: 140, loss: 0.1350616067647934
step: 150, loss: 0.024424416944384575
step: 160, loss: 0.030720949172973633
step: 170, loss: 0.019572535529732704
step: 180, loss: 0.16502612829208374
step: 190, loss: 0.08045855164527893
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7566137566137565, 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: 0.052802674472332
step: 10, loss: 0.014385807327926159
step: 20, loss: 0.1001448705792427
step: 30, loss: 0.0029828972183167934
step: 40, loss: 0.011339055374264717
step: 50, loss: 0.15522396564483643

```

```

step: 60, loss: 0.04547058790922165
step: 70, loss: 0.0027617602609097958
step: 80, loss: 0.009711699560284615
step: 90, loss: 0.05186249315738678
step: 100, loss: 0.03425687178969383
step: 110, loss: 0.0053167506121098995
step: 120, loss: 0.012436002492904663
step: 130, loss: 0.022917406633496284
step: 140, loss: 0.005588207859545946
step: 150, loss: 0.007968227379024029
step: 160, loss: 0.006884336471557617
step: 170, loss: 0.08656869828701019
step: 180, loss: 0.19017192721366882
step: 190, loss: 0.12800830602645874
/usr/local/lib/python3.7/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=0.7322404371584699, f1=0.747191011235955, 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: 0.004628111608326435
step: 10, loss: 0.002874290104955435
step: 20, loss: 0.007903682999312878
step: 30, loss: 0.0012711833696812391
step: 40, loss: 0.1112915500998497
step: 50, loss: 0.003551681060343981
step: 60, loss: 0.001150324009358883
step: 70, loss: 0.14062753319740295
step: 80, loss: 0.12659987807273865
step: 90, loss: 0.033522315323352814
step: 100, loss: 0.1459076851606369
step: 110, loss: 0.005344597157090902
step: 120, loss: 0.016978973522782326
step: 130, loss: 0.006231789942830801
step: 140, loss: 0.0059824250638484955
step: 150, loss: 0.009000658057630062
step: 160, loss: 0.024913690984249115
step: 170, loss: 0.012193228118121624
step: 180, loss: 0.010035126470029354
step: 190, loss: 0.006088257301598787
/usr/local/lib/python3.7/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=0.7314285714285714, f1=0.7420289855072463, 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: 0.00735838059335947
step: 10, loss: 0.00500513706356287
step: 20, loss: 0.06259159743785858
step: 30, loss: 0.015714822337031364
step: 40, loss: 0.021223634481430054
step: 50, loss: 0.04301673173904419
step: 60, loss: 0.006188428029417992
step: 70, loss: 0.00065319036366418
step: 80, loss: 0.03217722102999687
step: 90, loss: 0.001352612511254847
step: 100, loss: 0.008759778924286366
step: 110, loss: 0.0034356852993369102
step: 120, loss: 0.014568792656064034
step: 130, loss: 0.001336862682364881
step: 140, loss: 0.0021870448254048824
step: 150, loss: 0.12997940182685852
step: 160, loss: 0.0009081510943360627
step: 170, loss: 0.006095587275922298
step: 180, loss: 0.00936178583651781
step: 190, loss: 0.027128245681524277
/usr/local/lib/python3.7/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=0.720626631853786, f1=0.699724517906336, 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: 0.0009165782248601317
step: 10, loss: 0.004416976124048233
step: 20, loss: 0.008954016491770744
step: 30, loss: 0.024004651233553886
step: 40, loss: 0.014899530448019505
step: 50, loss: 0.001315238419920206
step: 60, loss: 0.002631099196150899
step: 70, loss: 0.0116411242634058
step: 80, loss: 0.0696617066860199
step: 90, loss: 0.08326394110918045
step: 100, loss: 0.022800764068961143
step: 110, loss: 0.001585022546350956
step: 120, loss: 0.03286293148994446
step: 130, loss: 0.00392469298094511
step: 140, loss: 0.08445583283901215
step: 150, loss: 0.04957854747772217
step: 160, loss: 0.08870478719472885
step: 170, loss: 0.0027452795766294003
step: 180, loss: 0.010029218159615993
step: 190, loss: 0.008912370540201664
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7578947368421052, f1=0.7193460490463215,
best_f1=0.7193460490463215
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009334200876764953
step: 10, loss: 0.009344487451016903
step: 20, loss: 0.05687889829277992
step: 30, loss: 0.012106361798942089
step: 40, loss: 0.1508532166481018
step: 50, loss: 0.0020532123744487762
step: 60, loss: 0.001417512190528214
step: 70, loss: 0.059755921363830566
step: 80, loss: 0.0011741239577531815
step: 90, loss: 0.004535073880106211
step: 100, loss: 0.01585112139582634
step: 110, loss: 0.005560542922466993
step: 120, loss: 0.21265575289726257
step: 130, loss: 0.05014912411570549
step: 140, loss: 0.0005597443087026477
step: 150, loss: 0.038524605333805084
step: 160, loss: 0.0052863918244838715
step: 170, loss: 0.0011823464883491397
step: 180, loss: 0.23885351419448853
step: 190, loss: 0.00711966585367918
/usr/local/lib/python3.7/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=0.7222222222222223, f1=0.7352185089974292,
best_f1=0.7193460490463215
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0013296327088028193
step: 10, loss: 0.016773564741015434
step: 20, loss: 0.002951235743239522
step: 30, loss: 0.004926509223878384
step: 40, loss: 0.0007426689262501895
step: 50, loss: 0.05206774175167084
step: 60, loss: 0.009801731444895267
step: 70, loss: 0.002346133114770055
step: 80, loss: 0.003822480561211705

```

```

step: 90, loss: 0.0041619883850216866
step: 100, loss: 0.014783866703510284
step: 110, loss: 0.006392575800418854
step: 120, loss: 0.0013381908647716045
step: 130, loss: 0.004751875530928373
step: 140, loss: 0.0005861123790964484
step: 150, loss: 0.005956978537142277
step: 160, loss: 0.005013534799218178
step: 170, loss: 0.06745847314596176
step: 180, loss: 0.0006255003390833735
step: 190, loss: 0.0014223877806216478
/usr/local/lib/python3.7/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=0.7362924281984334, f1=0.7427055702917772,
best_f1=0.7193460490463215
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.004314998164772987
step: 10, loss: 0.005764218512922525
step: 20, loss: 0.01949838548898697
step: 30, loss: 0.006635607685893774
step: 40, loss: 0.000499087676871568
step: 50, loss: 0.0009582301136106253
step: 60, loss: 0.0007006266387179494
step: 70, loss: 0.00034058454912155867
step: 80, loss: 0.0004531140730250627
step: 90, loss: 0.013797948136925697
step: 100, loss: 0.0007779004517942667
step: 110, loss: 0.001025402918457985
step: 120, loss: 0.012593901716172695
step: 130, loss: 0.0013026815140619874
step: 140, loss: 0.00035023532109335065
step: 150, loss: 0.0003293062618467957
step: 160, loss: 0.002847240073606372
step: 170, loss: 0.005875591188669205
step: 180, loss: 0.0047162617556750774
step: 190, loss: 0.001347940880805254
/usr/local/lib/python3.7/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=0.7301587301587302, f1=0.738544474393531,
best_f1=0.7193460490463215
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.014649216085672379

```

```

step: 10, loss: 0.0051207104697823524
step: 20, loss: 0.0005499412072822452
step: 30, loss: 0.004458258859813213
step: 40, loss: 0.002373563591390848
step: 50, loss: 0.0023730245884507895
step: 60, loss: 0.0012721663806587458
step: 70, loss: 0.0013391238171607256
step: 80, loss: 0.0014525667065754533
step: 90, loss: 0.0005073578795418143
step: 100, loss: 0.002404098864644766
step: 110, loss: 0.00037640921073034406
step: 120, loss: 0.0007657254463993013
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 130, loss: 0.0007427866803482175
step: 140, loss: 0.10750268399715424
step: 150, loss: 0.0006430202047340572
step: 160, loss: 0.007115840911865234
step: 170, loss: 0.0010979896178469062
step: 180, loss: 0.0010974104516208172
step: 190, loss: 0.008625308983027935
/usr/local/lib/python3.7/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=0.7247191011235956, f1=0.7256637168141593,
best_f1=0.7193460490463215
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0024830566253513098
step: 10, loss: 0.0009327545994892716
step: 20, loss: 0.0021844275761395693
step: 30, loss: 0.0025562045630067587
step: 40, loss: 0.0009709721780382097
step: 50, loss: 0.08794540166854858
step: 60, loss: 0.016025125980377197
step: 70, loss: 0.002392996335402131
step: 80, loss: 0.0018767045112326741
step: 90, loss: 0.0068210940808057785
step: 100, loss: 0.08552148938179016
step: 110, loss: 0.001279007294215262
step: 120, loss: 0.000768603291362524
step: 130, loss: 0.015417098067700863
step: 140, loss: 0.0015222834190353751
step: 150, loss: 0.005031174048781395
step: 160, loss: 0.0010594859486445785
step: 170, loss: 0.0627526044845581
step: 180, loss: 0.012408309616148472
step: 190, loss: 0.001845984603278339

```

```

/usr/local/lib/python3.7/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=0.7346938775510203, f1=0.7506561679790026,
best_f1=0.7193460490463215
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.05198203772306442
step: 10, loss: 0.001583832548931241
step: 20, loss: 0.0013023681240156293
step: 30, loss: 0.003011354710906744
step: 40, loss: 0.0010423206258565187
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 8192.0
step: 50, loss: 0.024196529760956764
step: 60, loss: 0.003006687154993415
step: 70, loss: 0.0019901813939213753
step: 80, loss: 0.0006529234815388918
step: 90, loss: 0.0007659975090064108
step: 100, loss: 0.001287168124690652
step: 110, loss: 0.0007044750964269042
step: 120, loss: 0.0017778262263163924
step: 130, loss: 0.001084089744836092
step: 140, loss: 0.010050147771835327
step: 150, loss: 0.012037123553454876
step: 160, loss: 0.0007208166061900556
step: 170, loss: 0.0007643489516340196
step: 180, loss: 0.002073202980682254
step: 190, loss: 0.0004381176840979606
/usr/local/lib/python3.7/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=0.745308310991957, f1=0.745945945945946,
best_f1=0.7193460490463215
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.002376118442043662
step: 10, loss: 0.04009251296520233
step: 20, loss: 0.03907996788620949
step: 30, loss: 0.0009244375978596509
step: 40, loss: 0.0007876799209043384
step: 50, loss: 0.0014237695140764117
step: 60, loss: 0.0007332486566156149
step: 70, loss: 0.0007852123235352337
step: 80, loss: 0.0004605756257660687
step: 90, loss: 0.006815875880420208
step: 100, loss: 0.0026656666304916143

```



```

step: 110, loss: 0.003202583407983184
step: 120, loss: 0.011935630813241005
step: 130, loss: 0.024416420608758926
step: 140, loss: 0.09239816665649414
step: 150, loss: 0.0018057468114420772
step: 160, loss: 0.002222601091489196
step: 170, loss: 0.0008057107916101813
step: 180, loss: 0.002798131899908185
step: 190, loss: 0.0010560646187514067
/usr/local/lib/python3.7/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=0.744186046511628, f1=0.7513227513227512,
best_f1=0.7193460490463215
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
2049it [00:06, 295.76it/s]
load_f1 = 0.7052341597796142
real_f1 = 0.7052341597796142
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-

```

```
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
4398it [00:11, 374.69it/s]
```

4.2 DITTO TEXTUAL

4.2.1 Abt-Buy - Running the matcher

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

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

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

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

Defaults for this optimization level are:

```

enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8375685214996338
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 10, loss: 0.22011253237724304
step: 20, loss: 0.15974129736423492
step: 30, loss: 0.2235192507505417
step: 40, loss: 0.31609347462654114
step: 50, loss: 0.3760320842266083
step: 60, loss: 0.43541476130485535
step: 70, loss: 0.3127385675907135
step: 80, loss: 0.24677300453186035
step: 90, loss: 0.4045160114765167
step: 100, loss: 0.2298102229833603
step: 110, loss: 0.17413237690925598
step: 120, loss: 0.5454750657081604
step: 130, loss: 0.42129018902778625
step: 140, loss: 0.44721171259880066

```

```

step: 150, loss: 0.07640772312879562
step: 160, loss: 0.25409606099128723
step: 170, loss: 0.14589089155197144
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.6883720930232559, f1=0.6914153132250581,
best_f1=0.6914153132250581
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.17728303372859955
step: 10, loss: 0.08009377866983414
step: 20, loss: 0.2419237196445465
step: 30, loss: 0.10363567620515823
step: 40, loss: 0.15749022364616394
step: 50, loss: 0.17200598120689392
step: 60, loss: 0.08957061171531677
step: 70, loss: 0.2312779277563095
step: 80, loss: 0.16389940679073334
step: 90, loss: 0.31140148639678955
step: 100, loss: 0.1458575576543808
step: 110, loss: 0.17919452488422394
step: 120, loss: 0.048248041421175
step: 130, loss: 0.0457073450088501
step: 140, loss: 0.1755364090204239
step: 150, loss: 0.12642623484134674
step: 160, loss: 0.10852645337581635
step: 170, loss: 0.1804024875164032
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.7352185089974292, f1=0.7549999999999999,
best_f1=0.7549999999999999
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07054009288549423
step: 10, loss: 0.02908700704574585
step: 20, loss: 0.03476757928729057

```

```

step: 30, loss: 0.23027083277702332
step: 40, loss: 0.02163136750459671
step: 50, loss: 0.14435505867004395
step: 60, loss: 0.07368088513612747
step: 70, loss: 0.04432620480656624
step: 80, loss: 0.29061242938041687
step: 90, loss: 0.07343100756406784
step: 100, loss: 0.023847905918955803
step: 110, loss: 0.055907584726810455
step: 120, loss: 0.013073399662971497
step: 130, loss: 0.12882404029369354
step: 140, loss: 0.007644230499863625
step: 150, loss: 0.12352359294891357
step: 160, loss: 0.021899884566664696
step: 170, loss: 0.19153250753879547
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.795969773299748, f1=0.7777777777777778,
best_f1=0.7777777777777778
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07947802543640137
step: 10, loss: 0.21320199966430664
step: 20, loss: 0.04353250563144684
step: 30, loss: 0.026810063049197197
step: 40, loss: 0.20865844190120697
step: 50, loss: 0.01231934130191803
step: 60, loss: 0.06942149996757507
step: 70, loss: 0.02227652631700039
step: 80, loss: 0.17072515189647675
step: 90, loss: 0.05175894871354103
step: 100, loss: 0.011254177428781986
step: 110, loss: 0.01390135195106268
step: 120, loss: 0.16997428238391876
step: 130, loss: 0.06576304882764816
step: 140, loss: 0.08768488466739655
step: 150, loss: 0.07447771728038788
step: 160, loss: 0.06418835371732712
step: 170, loss: 0.1307312250137329
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")

```

```

/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8184143222506394, f1=0.8, best_f1=0.8
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.02146870642900467
step: 10, loss: 0.01350027322769165
step: 20, loss: 0.00931326113641262
step: 30, loss: 0.055572520941495895
step: 40, loss: 0.014482787810266018
step: 50, loss: 0.13722877204418182
step: 60, loss: 0.0005929091130383313
step: 70, loss: 0.020907897502183914
step: 80, loss: 0.008356893435120583
step: 90, loss: 0.11456067860126495
step: 100, loss: 0.0764215886592865
step: 110, loss: 0.12772363424301147
step: 120, loss: 0.049246735870838165
step: 130, loss: 0.02941291406750679
step: 140, loss: 0.07519997656345367
step: 150, loss: 0.009557869285345078
step: 160, loss: 0.010325903072953224
step: 170, loss: 0.06722235679626465
/usr/local/lib/python3.7/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=0.8103896103896103, f1=0.8, best_f1=0.8
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07397440075874329
step: 10, loss: 0.0594138503074646
step: 20, loss: 0.010959993116557598
step: 30, loss: 0.029498843476176262
step: 40, loss: 0.0022445630747824907
step: 50, loss: 0.062418125569820404
step: 60, loss: 0.11785972118377686
step: 70, loss: 0.014430905692279339
step: 80, loss: 0.07258039712905884
step: 90, loss: 0.027674805372953415
step: 100, loss: 0.026813391596078873
step: 110, loss: 0.03609691932797432
step: 120, loss: 0.036605674773454666
step: 130, loss: 0.14400823414325714
step: 140, loss: 0.02162606082856655

```

```

step: 150, loss: 0.27811184525489807
step: 160, loss: 0.04500763863325119
step: 170, loss: 0.004288089461624622
/usr/local/lib/python3.7/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=0.8099999999999999, f1=0.7980769230769231, best_f1=0.8
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0015002090949565172
step: 10, loss: 0.026575816795229912
step: 20, loss: 0.005459794774651527
step: 30, loss: 0.010562906041741371
step: 40, loss: 0.011736735701560974
step: 50, loss: 0.01029485184699297
step: 60, loss: 0.11696487665176392
step: 70, loss: 0.02779070846736431
step: 80, loss: 0.06901451200246811
step: 90, loss: 0.06810961663722992
step: 100, loss: 0.007425442337989807
step: 110, loss: 0.07202039659023285
step: 120, loss: 0.1889137178659439
step: 130, loss: 0.17643390595912933
step: 140, loss: 0.015924254432320595
step: 150, loss: 0.036045704036951065
step: 160, loss: 0.019712986424565315
step: 170, loss: 0.1258598119020462
/usr/local/lib/python3.7/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=0.8124999999999999, f1=0.8044943820224719, best_f1=0.8
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.020575296133756638
step: 10, loss: 0.02990705333650112
step: 20, loss: 0.007721650414168835
step: 30, loss: 0.014930998906493187
step: 40, loss: 0.00285227969288826
step: 50, loss: 0.004807566292583942
step: 60, loss: 0.0046861013397574425
step: 70, loss: 0.0204416885972023
step: 80, loss: 0.006495464593172073
step: 90, loss: 0.002079028869047761
step: 100, loss: 0.005124358460307121
step: 110, loss: 0.03067142330110073
step: 120, loss: 0.16243059933185577

```

```

step: 130, loss: 0.002544563729315996
step: 140, loss: 0.009559041820466518
step: 150, loss: 0.10647299885749817
step: 160, loss: 0.014301163144409657
step: 170, loss: 0.0029167335014790297
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8260869565217391, f1=0.8, best_f1=0.8
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003031868487596512
step: 10, loss: 0.13832782208919525
step: 20, loss: 0.007887877523899078
step: 30, loss: 0.03427259996533394
step: 40, loss: 0.04290604963898659
step: 50, loss: 0.12757883965969086
step: 60, loss: 0.009918943047523499
step: 70, loss: 0.006238831672817469
step: 80, loss: 0.010359224863350391
step: 90, loss: 0.018563518300652504
step: 100, loss: 0.01208871603012085
step: 110, loss: 0.03078300878405571
step: 120, loss: 0.009360991418361664
step: 130, loss: 0.0006976443692110479
step: 140, loss: 0.003986900206655264
step: 150, loss: 0.020223239436745644
step: 160, loss: 0.04558350518345833
step: 170, loss: 0.014535272493958473
/usr/local/lib/python3.7/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=0.8238213399503721, f1=0.7990654205607476, best_f1=0.8
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.009699133224785328
step: 10, loss: 0.034731388092041016
step: 20, loss: 0.0009268664289265871
step: 30, loss: 0.0006466585909947753
step: 40, loss: 0.10909643769264221
step: 50, loss: 0.04094359278678894
step: 60, loss: 0.04030206426978111

```



```

step: 70, loss: 0.0006100141326896846
step: 80, loss: 0.003421430941671133
step: 90, loss: 0.04597768187522888
step: 100, loss: 0.04265829920768738
step: 110, loss: 0.0018275165930390358
step: 120, loss: 0.039224766194820404
step: 130, loss: 0.11584732681512833
step: 140, loss: 0.012396431528031826
step: 150, loss: 0.004260716028511524
step: 160, loss: 0.0019327938789501786
step: 170, loss: 0.019669657573103905
/usr/local/lib/python3.7/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=0.824742268041237, f1=0.8048780487804877, best_f1=0.8
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0003571540000848472
step: 10, loss: 0.016728980466723442
step: 20, loss: 0.06923864036798477
step: 30, loss: 0.003718274412676692
step: 40, loss: 0.025335414335131645
step: 50, loss: 0.002963831415399909
step: 60, loss: 0.010091601870954037
step: 70, loss: 0.0015825120499357581
step: 80, loss: 0.0025856359861791134
step: 90, loss: 0.0036495935637503862
step: 100, loss: 0.0055188084952533245
step: 110, loss: 0.00465945852920413
step: 120, loss: 0.003157163504511118
step: 130, loss: 0.000982215628027916
step: 140, loss: 0.053159911185503006
step: 150, loss: 0.004760757088661194
step: 160, loss: 0.005075538065284491
step: 170, loss: 0.1475149393081665
/usr/local/lib/python3.7/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=0.8226221079691517, f1=0.7970660146699265, best_f1=0.8
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008237511850893497
step: 10, loss: 0.004081185441464186
step: 20, loss: 0.01505288202315569
step: 30, loss: 0.0003635978209786117
step: 40, loss: 0.000981884659267962

```

```

step: 50, loss: 0.009909273125231266
step: 60, loss: 0.0002493297215551138
step: 70, loss: 0.015140803530812263
step: 80, loss: 0.058578141033649445
step: 90, loss: 0.000717545161023736
step: 100, loss: 0.010572141036391258
step: 110, loss: 0.00021319107327144593
step: 120, loss: 0.0038952503819018602
step: 130, loss: 0.05155014619231224
step: 140, loss: 0.009475287050008774
step: 150, loss: 0.0024653999134898186
step: 160, loss: 0.003593552391976118
step: 170, loss: 0.00025815056869760156
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8337468982630273, f1=0.7971360381861575,
best_f1=0.7971360381861575
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.003963058348745108
step: 10, loss: 0.06640563905239105
step: 20, loss: 0.00047251128125935793
step: 30, loss: 0.0009522044565528631
step: 40, loss: 0.01218385435640812
step: 50, loss: 0.01974751241505146
step: 60, loss: 0.0007981158560141921
step: 70, loss: 0.014183172024786472
step: 80, loss: 0.0012827663449570537
step: 90, loss: 0.0041090562008321285
step: 100, loss: 0.0009316182695329189
step: 110, loss: 0.002317501697689295
step: 120, loss: 0.000724312209058553
step: 130, loss: 0.004678455181419849
step: 140, loss: 0.0009547442314215004
step: 150, loss: 0.0008178868447430432
step: 160, loss: 0.006839803885668516
step: 170, loss: 0.0021106272470206022
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.8385542168674698, f1=0.7954545454545455,
best_f1=0.7954545454545455
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0014160298742353916
step: 10, loss: 0.00027595399296842515
step: 20, loss: 0.009776494465768337
step: 30, loss: 0.00036931553040631115
step: 40, loss: 0.00045950888306833804
step: 50, loss: 0.020281502977013588
step: 60, loss: 0.0002601188898552209
step: 70, loss: 0.0050580669194459915
step: 80, loss: 0.002726594917476177
step: 90, loss: 0.000531783327460289
step: 100, loss: 0.005510891322046518
step: 110, loss: 0.006377338897436857
step: 120, loss: 0.002280455082654953
step: 130, loss: 0.0010302424198016524
step: 140, loss: 0.0007509610150009394
step: 150, loss: 0.008680530823767185
step: 160, loss: 0.1849803626537323
step: 170, loss: 0.0009624166414141655
/usr/local/lib/python3.7/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=0.8296296296296296, f1=0.8047058823529412,
best_f1=0.7954545454545455
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.001595101086422801
step: 10, loss: 0.0062105972319841385
step: 20, loss: 0.12639786303043365
step: 30, loss: 0.0048089041374623775
step: 40, loss: 0.0008235956775024533
step: 50, loss: 0.0138109615072608
step: 60, loss: 0.00029089904273860157
step: 70, loss: 0.0018249659333378077
step: 80, loss: 0.002152160042896867
step: 90, loss: 0.0029170450288802385
step: 100, loss: 0.0017596352845430374
step: 110, loss: 0.0003988875832874328
step: 120, loss: 0.029041927307844162
step: 130, loss: 0.0007864783983677626
step: 140, loss: 0.000983992125838995

```

```

step: 150, loss: 0.06848878413438797
step: 160, loss: 0.00021883394219912589
step: 170, loss: 0.005524058360606432
/usr/local/lib/python3.7/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=0.834567901234568, f1=0.8047058823529412,
best_f1=0.7954545454545455
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02: FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
1916it [00:04, 429.29it/s]
load_f1 = 0.6219512195121952
real_f1 = 0.5938697318007663
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:12, 355.67it/s]

```

4.3 DITTO DIRTY

4.3.1 DBLP-ACM - Running the matcher

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

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

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

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

Defaults for this optimization level are:

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

```

master_weights      : True
loss_scale          : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled              : True
opt_level            : 02
cast_model_type      : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32   : True
master_weights       : True
loss_scale           : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.8068530559539795
step: 10, loss: 0.46482306718826294
step: 20, loss: 0.5797284245491028
step: 30, loss: 0.4295070469379425
step: 40, loss: 0.1695743352174759
step: 50, loss: 0.06050645187497139
step: 60, loss: 0.17647603154182434
step: 70, loss: 0.12155285477638245
step: 80, loss: 0.20841144025325775
step: 90, loss: 0.10025109350681305
step: 100, loss: 0.14084477722644806
step: 110, loss: 0.055201757699251175
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 120, loss: 0.015062828548252583
step: 130, loss: 0.023821523413062096
step: 140, loss: 0.03638242557644844
step: 150, loss: 0.129644513130188
step: 160, loss: 0.26726034283638
step: 170, loss: 0.07824092358350754
step: 180, loss: 0.03351068124175072
step: 190, loss: 0.026103954762220383

```

```

step: 200, loss: 0.004206375684589148
step: 210, loss: 0.017186248674988747
step: 220, loss: 0.008755041286349297
step: 230, loss: 0.05393950268626213
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9636163175303197, f1=0.96875, best_f1=0.96875
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.036267075687646866
step: 10, loss: 0.022117028012871742
step: 20, loss: 0.00925963744521141
step: 30, loss: 0.008652998134493828
step: 40, loss: 0.021740414202213287
step: 50, loss: 0.016606532037258148
step: 60, loss: 0.00620679697021842
step: 70, loss: 0.009339447133243084
step: 80, loss: 0.018406907096505165
step: 90, loss: 0.013423118740320206
step: 100, loss: 0.12303227931261063
step: 110, loss: 0.053937915712594986
step: 120, loss: 0.038175325840711594
step: 130, loss: 0.013877375982701778
step: 140, loss: 0.3640338182449341
step: 150, loss: 0.017322445288300514
step: 160, loss: 0.01657184399664402
step: 170, loss: 0.08425473421812057
step: 180, loss: 0.00947324838489294
step: 190, loss: 0.11024720221757889
step: 200, loss: 0.010807884857058525
step: 210, loss: 0.15070652961730957
step: 220, loss: 0.0029593599028885365
step: 230, loss: 0.01476271077990532
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9733924611973392, f1=0.9655937846836848,
best_f1=0.9655937846836848

```

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.1432190090417862
step: 10, loss: 0.03255777433514595
step: 20, loss: 0.004005769267678261
step: 30, loss: 0.013311421498656273
step: 40, loss: 0.10384971648454666
step: 50, loss: 0.012228749692440033
step: 60, loss: 0.009727206081151962
step: 70, loss: 0.02792849764227867
step: 80, loss: 0.0063607716001570225
step: 90, loss: 0.12220769375562668
step: 100, loss: 0.005493273492902517
step: 110, loss: 0.024107465520501137
step: 120, loss: 0.0038526025600731373
step: 130, loss: 0.000758830108679831
step: 140, loss: 0.0031503071077167988
step: 150, loss: 0.002187041100114584
step: 160, loss: 0.002501563634723425
step: 170, loss: 0.0890793725848198
step: 180, loss: 0.004308342933654785
step: 190, loss: 0.023301251232624054
step: 200, loss: 0.005488370545208454
step: 210, loss: 0.005554764531552792
step: 220, loss: 0.007536514662206173
step: 230, loss: 0.07103168964385986
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9765363128491621, f1=0.9708520179372198,
best_f1=0.9708520179372198
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.008017138577997684
step: 10, loss: 0.0031231867615133524
step: 20, loss: 0.0011198039865121245
step: 30, loss: 0.006072239484637976
step: 40, loss: 0.008991909213364124
step: 50, loss: 0.001742913736961782
step: 60, loss: 0.007258742116391659
step: 70, loss: 0.030726438388228416
step: 80, loss: 0.08021632581949234

```



```

step: 90, loss: 0.0017441310919821262
step: 100, loss: 0.0015435778768733144
step: 110, loss: 0.010357947088778019
step: 120, loss: 0.006193777546286583
step: 130, loss: 0.0016471416456624866
step: 140, loss: 0.007075991481542587
step: 150, loss: 0.007476191036403179
step: 160, loss: 0.018146319314837456
step: 170, loss: 0.013329719193279743
step: 180, loss: 0.08326032757759094
step: 190, loss: 0.009861325845122337
step: 200, loss: 0.01820707693696022
step: 210, loss: 0.0006135236471891403
step: 220, loss: 0.0016053662402555346
step: 230, loss: 0.0005897234077565372
/usr/local/lib/python3.7/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=0.9730941704035874, f1=0.9696969696969697,
best_f1=0.9708520179372198
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0009672529413364828
step: 10, loss: 0.0026438753120601177
step: 20, loss: 0.004170191008597612
step: 30, loss: 0.00040789193008095026
step: 40, loss: 0.0004935578326694667
step: 50, loss: 0.005628598853945732
step: 60, loss: 0.001147955423220992
step: 70, loss: 0.00018399243708699942
step: 80, loss: 0.008369057439267635
step: 90, loss: 0.0006059366860426962
step: 100, loss: 0.0072179134003818035
step: 110, loss: 0.002803327515721321
step: 120, loss: 0.022487415000796318
step: 130, loss: 0.09713273495435715
step: 140, loss: 0.0010837294394150376
step: 150, loss: 0.0006595014710910618
step: 160, loss: 0.11596768349409103
step: 170, loss: 0.11811453104019165
step: 180, loss: 0.00254966807551682
step: 190, loss: 0.013172395527362823
step: 200, loss: 0.006539487279951572
step: 210, loss: 0.0018142500193789601
step: 220, loss: 0.0017340565100312233
step: 230, loss: 0.00536087341606617
/usr/local/lib/python3.7/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=0.9722530521642618,
best_f1=0.9708520179372198
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0045935423113405704
step: 10, loss: 0.002691438188776374
step: 20, loss: 0.0009380376432090998
step: 30, loss: 0.0004857177846133709
step: 40, loss: 0.005821607541292906
step: 50, loss: 0.06571042537689209
step: 60, loss: 0.002968459390103817
step: 70, loss: 0.00853948574513197
step: 80, loss: 0.003517675679177046
step: 90, loss: 0.0003356283123139292
step: 100, loss: 0.00040778476977720857
step: 110, loss: 0.0423857606947422
step: 120, loss: 0.0005429275333881378
step: 130, loss: 0.0009267654386349022
step: 140, loss: 0.001127016730606556
step: 150, loss: 0.0008527192403562367
step: 160, loss: 0.0006013012607581913
step: 170, loss: 0.00032340845791622996
step: 180, loss: 0.0014649925287812948
step: 190, loss: 0.0561550073325634
step: 200, loss: 0.02422744780778885
step: 210, loss: 0.00321209873072803
step: 220, loss: 0.0008370807627215981
step: 230, loss: 0.003109723562374711
/usr/local/lib/python3.7/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=0.9753914988814317, f1=0.9675977653631285,
best_f1=0.9708520179372198
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.16177675127983093
step: 10, loss: 0.0014114024816080928
step: 20, loss: 0.00411283178254962
step: 30, loss: 0.0034620268270373344
step: 40, loss: 0.05191018804907799
step: 50, loss: 0.0003502040635794401
step: 60, loss: 0.00864966306835413
step: 70, loss: 0.0010212655179202557
step: 80, loss: 0.0002595896949060261

```

```

step: 90, loss: 0.012940850108861923
step: 100, loss: 0.003672546474263072
step: 110, loss: 0.0029038856737315655
step: 120, loss: 0.01717103272676468
step: 130, loss: 0.001457523088902235
step: 140, loss: 0.00019200103997718543
step: 150, loss: 0.005986642558127642
step: 160, loss: 0.0005236329161562026
step: 170, loss: 0.00019754754612222314
step: 180, loss: 0.00014400244981516153
step: 190, loss: 0.032296065241098404
step: 200, loss: 0.001530586974695325
step: 210, loss: 0.00022848336084280163
step: 220, loss: 0.006552666891366243
step: 230, loss: 0.04547060653567314
/usr/local/lib/python3.7/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=0.9689578713968958, f1=0.9709821428571428,
best_f1=0.9708520179372198
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.012154946103692055
step: 10, loss: 0.024474043399095535
step: 20, loss: 0.0013833114644512534
step: 30, loss: 0.00024757831124588847
step: 40, loss: 0.06483633816242218
step: 50, loss: 0.017136331647634506
step: 60, loss: 0.00022975224419496953
step: 70, loss: 0.0003725741698872298
step: 80, loss: 0.00038369439425878227
step: 90, loss: 0.059437040239572525
step: 100, loss: 0.00016588786093052477
step: 110, loss: 0.00025059416657313704
step: 120, loss: 0.00012263184180483222
step: 130, loss: 0.020171578973531723
step: 140, loss: 9.924564074026421e-05
step: 150, loss: 0.0006670156726613641
step: 160, loss: 0.001084994408302009
step: 170, loss: 0.0010011474369093776
step: 180, loss: 0.012542792595922947
step: 190, loss: 0.0047564213164150715
step: 200, loss: 0.011783869937062263
step: 210, loss: 0.00030410970794036984
step: 220, loss: 0.005137776955962181
step: 230, loss: 0.02965865284204483
/usr/local/lib/python3.7/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=0.9753914988814317, f1=0.9730941704035874,
best_f1=0.9708520179372198
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.005528390407562256
step: 10, loss: 0.001447893911972642
step: 20, loss: 0.0008693624986335635
step: 30, loss: 0.00024646453675813973
step: 40, loss: 9.190927812596783e-05
step: 50, loss: 0.0004331287636887282
step: 60, loss: 0.00017313948774244636
step: 70, loss: 0.028976544737815857
step: 80, loss: 0.03485580161213875
step: 90, loss: 0.017072705551981926
step: 100, loss: 0.0009356778464280069
step: 110, loss: 0.007913805544376373
step: 120, loss: 0.04799611493945122
step: 130, loss: 0.0001391739642713219
step: 140, loss: 0.04659472033381462
step: 150, loss: 0.0003090953396167606
step: 160, loss: 0.0005055581568740308
step: 170, loss: 0.00532831996679306
step: 180, loss: 0.0016778477001935244
step: 190, loss: 0.00022696866653859615
step: 200, loss: 0.0006143806967884302
step: 210, loss: 0.00010222003766102716
step: 220, loss: 0.0444820411503315
step: 230, loss: 0.0013663143618032336
/usr/local/lib/python3.7/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=0.9764837625979844, f1=0.9730337078651685,
best_f1=0.9708520179372198
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00010899777407757938
step: 10, loss: 0.00024273160670418292
step: 20, loss: 0.0012594213476404548
step: 30, loss: 0.00018251767323818058
step: 40, loss: 8.949984476203099e-05
step: 50, loss: 0.060849957168102264
step: 60, loss: 0.08211153000593185
step: 70, loss: 0.0011113706277683377
step: 80, loss: 0.0011342349462211132

```

```

step: 90, loss: 0.00025640460080467165
step: 100, loss: 0.0005436728824861348
step: 110, loss: 0.0003041523159481585
step: 120, loss: 0.012837840244174004
step: 130, loss: 0.007706218399107456
step: 140, loss: 0.021267537027597427
step: 150, loss: 0.00017226119234692305
step: 160, loss: 0.0010466292733326554
step: 170, loss: 0.00021757314971182495
step: 180, loss: 0.04761894792318344
step: 190, loss: 0.02551417611539364
step: 200, loss: 0.0119852339848876
step: 210, loss: 0.00016585239791311324
step: 220, loss: 0.00027042641886509955
step: 230, loss: 0.0013634171336889267
/usr/local/lib/python3.7/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=0.9763779527559054, f1=0.9707865168539327,
best_f1=0.9708520179372198
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00023301241162698716
step: 10, loss: 0.0013035458978265524
step: 20, loss: 7.666215969948098e-05
step: 30, loss: 0.00017108792962972075
step: 40, loss: 0.023873252794146538
step: 50, loss: 0.015489054843783379
step: 60, loss: 0.004264179151505232
step: 70, loss: 0.012224842794239521
step: 80, loss: 0.0003451011434663087
step: 90, loss: 0.0017265417845919728
step: 100, loss: 0.00012620282359421253
step: 110, loss: 0.002811027690768242
step: 120, loss: 0.0003331585612613708
step: 130, loss: 5.066504672868177e-05
step: 140, loss: 8.135532698361203e-05
step: 150, loss: 5.42332163604442e-05
step: 160, loss: 0.0001566020946484059
step: 170, loss: 0.002418744144961238
step: 180, loss: 0.0004607352311722934
step: 190, loss: 0.0005630454397760332
step: 200, loss: 0.0012771005276590586
step: 210, loss: 0.0016004128847271204
step: 220, loss: 0.00011010150046786293
step: 230, loss: 0.007551465183496475
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:

```

```

An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9787709497206705, f1=0.9742441209406495,
best_f1=0.9742441209406495
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00012358104868326336
step: 10, loss: 0.018774421885609627
step: 20, loss: 6.405053864000365e-05
step: 30, loss: 0.0009310066816397011
step: 40, loss: 0.00012372626224532723
step: 50, loss: 0.00028139824280515313
step: 60, loss: 0.016774982213974
step: 70, loss: 0.002329649403691292
step: 80, loss: 0.00020458397921174765
step: 90, loss: 0.0010396301513537765
step: 100, loss: 0.00019186994177289307
step: 110, loss: 0.00032750130048952997
step: 120, loss: 0.000209847537917085
step: 130, loss: 0.00027269148267805576
step: 140, loss: 9.547868103254586e-05
step: 150, loss: 0.0009357991511933506
step: 160, loss: 0.0017132613575085998
step: 170, loss: 0.00014060856483411044
step: 180, loss: 0.00010033689613919705
step: 190, loss: 0.0013039412442594767
step: 200, loss: 0.004595798905938864
step: 210, loss: 0.0002897736558225006
step: 220, loss: 0.023063527420163155
step: 230, loss: 0.00016373032121919096
/usr/local/lib/python3.7/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=0.9774774774774775, f1=0.9752808988764046,
best_f1=0.9742441209406495
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00015564152272418141
step: 10, loss: 0.000834843551274389
step: 20, loss: 0.00013318844139575958
step: 30, loss: 0.0041467053815722466
step: 40, loss: 0.0006063762702979147

```

```

step: 50, loss: 0.00016717438120394945
step: 60, loss: 0.00032941572135314345
step: 70, loss: 0.000148466628161259
step: 80, loss: 0.00013054565351922065
step: 90, loss: 9.483257599640638e-05
step: 100, loss: 0.0002444878628011793
step: 110, loss: 0.00010554141772445291
step: 120, loss: 0.0013371349778026342
step: 130, loss: 0.00017039172234945
step: 140, loss: 0.039784759283065796
step: 150, loss: 0.07893070578575134
step: 160, loss: 8.797438931651413e-05
step: 170, loss: 0.0002319986670045182
step: 180, loss: 0.0004702077421825379
step: 190, loss: 7.722299051238224e-05
step: 200, loss: 0.0001567625004099682
step: 210, loss: 0.00018382789858151227
step: 220, loss: 0.00019409548258408904
step: 230, loss: 0.0005397022468969226
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9797752808988766, f1=0.9753363228699552,
best_f1=0.9753363228699552
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 9.188614785671234e-05
step: 10, loss: 4.138202712056227e-05
step: 20, loss: 0.002727187005802989
step: 30, loss: 0.0030247231479734182
step: 40, loss: 4.839950634050183e-05
step: 50, loss: 9.233851596945897e-05
step: 60, loss: 0.0003660544170998037
step: 70, loss: 9.716938075143844e-05
step: 80, loss: 0.0002177508285967633
step: 90, loss: 0.00037102363421581686
step: 100, loss: 0.00705319456756115
step: 110, loss: 0.0001245046005351469
step: 120, loss: 0.0003119300527032465
step: 130, loss: 7.068615377647802e-05
step: 140, loss: 7.777327118674293e-05
step: 150, loss: 0.0003124246431980282
step: 160, loss: 4.328150680521503e-05

```

```

step: 170, loss: 0.00014773162547498941
step: 180, loss: 0.00011870460002683103
step: 190, loss: 0.0027262838557362556
step: 200, loss: 8.2820923125837e-05
step: 210, loss: 0.010572866536676884
step: 220, loss: 0.00017196797125507146
step: 230, loss: 5.2392475481610745e-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=0.9775784753363228, f1=0.9742441209406495,
best_f1=0.9753363228699552
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.005112274782732e-05
step: 10, loss: 0.0002767084224615246
step: 20, loss: 9.763568232301623e-05
step: 30, loss: 0.00020277539442759007
step: 40, loss: 9.410828351974487e-05
step: 50, loss: 0.00021424416627269238
step: 60, loss: 7.126632408471778e-05
step: 70, loss: 0.00014824382378719747
step: 80, loss: 0.00022294621157925576
step: 90, loss: 3.4863700420828536e-05
step: 100, loss: 7.661111885681748e-05
step: 110, loss: 0.0001585559220984578
step: 120, loss: 0.0001272447989322245
step: 130, loss: 0.0002444678684696555
step: 140, loss: 0.018984781578183174
step: 150, loss: 0.00024869211483746767
step: 160, loss: 0.0006184166413731873
step: 170, loss: 4.555821215035394e-05
step: 180, loss: 6.872993981232867e-05
step: 190, loss: 0.0052426462061703205
step: 200, loss: 4.813601481146179e-05
step: 210, loss: 0.04717394709587097
step: 220, loss: 0.011998492293059826
step: 230, loss: 8.301834168378264e-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=0.978675645342312, f1=0.9741863075196409,
best_f1=0.9753363228699552
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level O2:  FP16 training with FP32 batchnorm and FP32
master weights.

```



```

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions  : False
keep_batchnorm_fp32    : True
master_weights         : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
0it [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
2473it [00:08, 306.40it/s]
load_f1 = 0.9755011135857461
real_f1 = 0.9744160177975528
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, 381.08it/s]

```

4.3.2 DBLP-GoogleScholar - Running the matcher

```

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

```

```

--finetuning \
--lm distilbert \
--fp16 \
--save_model \
--dk general \
--dk product \
--summarize \
--da entry_swap \
--da attr_del

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

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

```

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

Defaults for this optimization level are:

```

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

```

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

After processing overrides, optimization options are:

```

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

```

Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly because apex was installed without --cuda_ext --cpp_ext. Using Python fallback. Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")

```

/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 32768.0
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:136:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order:
`optimizer.step()` before `lr_scheduler.step()`. Failure to do this will result
in PyTorch skipping the first value of the learning rate schedule. See more
details at https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-
rate
    "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate",
UserWarning)
step: 0, loss: 0.7912074327468872
step: 10, loss: 0.42237988114356995
step: 20, loss: 0.4963975250720978
step: 30, loss: 0.428940087556839
step: 40, loss: 0.3192249536514282
step: 50, loss: 0.24435310065746307
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 60, loss: 0.17945577204227448
step: 70, loss: 0.1405685991048813
step: 80, loss: 0.07005196809768677
step: 90, loss: 0.09204931557178497
step: 100, loss: 0.32400691509246826
step: 110, loss: 0.12258554995059967
step: 120, loss: 0.03670407459139824
step: 130, loss: 0.021633747965097427
step: 140, loss: 0.19450409710407257
step: 150, loss: 0.10079333186149597
step: 160, loss: 0.1571348011493683
step: 170, loss: 0.1494596004486084
step: 180, loss: 0.11447524279356003
step: 190, loss: 0.12027208507061005
step: 200, loss: 0.11355118453502655
step: 210, loss: 0.1484166532754898
step: 220, loss: 0.1034451425075531
step: 230, loss: 0.10695800185203552
step: 240, loss: 0.13235707581043243
step: 250, loss: 0.06053410843014717
step: 260, loss: 0.023847095668315887
step: 270, loss: 0.06642293930053711
step: 280, loss: 0.19405631721019745
step: 290, loss: 0.023907694965600967
step: 300, loss: 0.2456458956003189
step: 310, loss: 0.10391848534345627
step: 320, loss: 0.16537891328334808
step: 330, loss: 0.14243292808532715

```

```

step: 340, loss: 0.13500116765499115
step: 350, loss: 0.14577339589595795
step: 360, loss: 0.11467761546373367
step: 370, loss: 0.15530531108379364
step: 380, loss: 0.22132271528244019
step: 390, loss: 0.02539343759417534
step: 400, loss: 0.007706941571086645
step: 410, loss: 0.028480513021349907
step: 420, loss: 0.025601115077733994
step: 430, loss: 0.05962274968624115
step: 440, loss: 0.1096535325050354
step: 450, loss: 0.08720386028289795
step: 460, loss: 0.17899034917354584
step: 470, loss: 0.2236630618572235
step: 480, loss: 0.26306119561195374
step: 490, loss: 0.04593901336193085
step: 500, loss: 0.022532906383275986
step: 510, loss: 0.08174385130405426
step: 520, loss: 0.06542421877384186
step: 530, loss: 0.09419811517000198
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.925497454881999, f1=0.9214417744916821,
best_f1=0.9214417744916821
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.07053380459547043
step: 10, loss: 0.19120557606220245
step: 20, loss: 0.18858741223812103
step: 30, loss: 0.18480953574180603
step: 40, loss: 0.012392549775540829
step: 50, loss: 0.042816005647182465
step: 60, loss: 0.15702597796916962
step: 70, loss: 0.28468841314315796
step: 80, loss: 0.026971209794282913
step: 90, loss: 0.004788169637322426
step: 100, loss: 0.19934721291065216
step: 110, loss: 0.042254216969013214
step: 120, loss: 0.06966745853424072
step: 130, loss: 0.02703266032040119
step: 140, loss: 0.07797323912382126
step: 150, loss: 0.04174526780843735

```

```
step: 160, loss: 0.01021086610853672
step: 170, loss: 0.1536986082792282
step: 180, loss: 0.006883625872433186
step: 190, loss: 0.10996340215206146
step: 200, loss: 0.03599987179040909
step: 210, loss: 0.030912213027477264
step: 220, loss: 0.17751003801822662
step: 230, loss: 0.05811040848493576
step: 240, loss: 0.1798432469367981
step: 250, loss: 0.07443498075008392
step: 260, loss: 0.016705410555005074
step: 270, loss: 0.37645843625068665
step: 280, loss: 0.29862886667251587
step: 290, loss: 0.16928809881210327
step: 300, loss: 0.1270899623632431
step: 310, loss: 0.07421943545341492
step: 320, loss: 0.2394046187400818
step: 330, loss: 0.07876625657081604
step: 340, loss: 0.004119782242923975
step: 350, loss: 0.04330924153327942
step: 360, loss: 0.026733752340078354
step: 370, loss: 0.0101504260674119
step: 380, loss: 0.061535030603408813
step: 390, loss: 0.029833350330591202
step: 400, loss: 0.023047881200909615
step: 410, loss: 0.002834440441802144
step: 420, loss: 0.04083476960659027
step: 430, loss: 0.018594039604067802
step: 440, loss: 0.009615584276616573
step: 450, loss: 0.05192630738019943
step: 460, loss: 0.3820788562297821
step: 470, loss: 0.14157091081142426
step: 480, loss: 0.24344995617866516
step: 490, loss: 0.054353393614292145
step: 500, loss: 0.03618433326482773
step: 510, loss: 0.020116528496146202
step: 520, loss: 0.100272037088871
step: 530, loss: 0.11634951829910278
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
/usr/local/lib/python3.7/dist-packages/torch/optim/lr_scheduler.py:216:
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=0.9351724137931035, f1=0.9267618608935975,
best_f1=0.9267618608935975
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
```

An input tensor was not cuda.

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

```
step: 0, loss: 0.007696039043366909
step: 10, loss: 0.14963561296463013
step: 20, loss: 0.1368931531906128
step: 30, loss: 0.21777202188968658
step: 40, loss: 0.010042200796306133
step: 50, loss: 0.03999880328774452
step: 60, loss: 0.0036386854480952024
step: 70, loss: 0.0801500529050827
step: 80, loss: 0.12551560997962952
step: 90, loss: 0.03479592502117157
step: 100, loss: 0.06403497606515884
step: 110, loss: 0.04390610754489899
step: 120, loss: 0.07039905339479446
step: 130, loss: 0.008658258244395256
step: 140, loss: 0.04834841564297676
step: 150, loss: 0.010034697130322456
step: 160, loss: 0.002243109978735447
step: 170, loss: 0.013784680515527725
step: 180, loss: 0.041471026837825775
step: 190, loss: 0.009797568432986736
step: 200, loss: 0.006336461286991835
step: 210, loss: 0.12520481646060944
step: 220, loss: 0.010764580219984055
step: 230, loss: 0.04136223345994949
step: 240, loss: 0.13521040976047516
step: 250, loss: 0.05761308595538139
step: 260, loss: 0.015916917473077774
step: 270, loss: 0.002840097527951002
step: 280, loss: 0.030244404450058937
step: 290, loss: 0.04103052616119385
step: 300, loss: 0.056507501751184464
step: 310, loss: 0.1750049740076065
step: 320, loss: 0.16804425418376923
step: 330, loss: 0.0052512818947434425
step: 340, loss: 0.04340745136141777
step: 350, loss: 0.004770034458488226
step: 360, loss: 0.008506901562213898
step: 370, loss: 0.007688440848141909
step: 380, loss: 0.02670394629240036
step: 390, loss: 0.03681650757789612
step: 400, loss: 0.09714610129594803
step: 410, loss: 0.020033396780490875
step: 420, loss: 0.06467558443546295
step: 430, loss: 0.020722782239317894
step: 440, loss: 0.058495502918958664
step: 450, loss: 0.011044413782656193
```

```
step: 460, loss: 0.04484616592526436
step: 470, loss: 0.011480595916509628
step: 480, loss: 0.11115462332963943
step: 490, loss: 0.0259197186678648
step: 500, loss: 0.15148723125457764
step: 510, loss: 0.0010629555908963084
step: 520, loss: 0.021790292114019394
step: 530, loss: 0.023784376680850983
/usr/local/lib/python3.7/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=0.9227188081936686, f1=0.9215686274509803,
best_f1=0.9267618608935975
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0040315305814146996
step: 10, loss: 0.022235361859202385
step: 20, loss: 0.014285593293607235
step: 30, loss: 0.009488536044955254
step: 40, loss: 0.03775034099817276
step: 50, loss: 0.026805086061358452
step: 60, loss: 0.06642568111419678
step: 70, loss: 0.0044234199449419975
step: 80, loss: 0.04726732149720192
step: 90, loss: 0.06625312566757202
step: 100, loss: 0.024724029004573822
step: 110, loss: 0.0012881840812042356
step: 120, loss: 0.038373664021492004
step: 130, loss: 0.018030015751719475
step: 140, loss: 0.05093259736895561
step: 150, loss: 0.0021981976460665464
step: 160, loss: 0.0008278798195533454
step: 170, loss: 0.002686135470867157
step: 180, loss: 0.013359637930989265
step: 190, loss: 0.03247581794857979
step: 200, loss: 0.006395548582077026
step: 210, loss: 0.032337486743927
step: 220, loss: 0.018741216510534286
step: 230, loss: 0.15471448004245758
step: 240, loss: 0.036876220256090164
step: 250, loss: 0.12238455563783646
step: 260, loss: 0.11521373689174652
step: 270, loss: 0.03222065418958664
step: 280, loss: 0.005336155649274588
step: 290, loss: 0.1263098418712616
step: 300, loss: 0.001794960000552237
step: 310, loss: 0.0045158653520047665
```

```

step: 320, loss: 0.006121852435171604
step: 330, loss: 0.06574640423059464
step: 340, loss: 0.010984521359205246
step: 350, loss: 0.022131381556391716
step: 360, loss: 0.014696259051561356
step: 370, loss: 0.02414068765938282
step: 380, loss: 0.004510974977165461
step: 390, loss: 0.01752157136797905
step: 400, loss: 0.001640588161535561
step: 410, loss: 0.007222972344607115
step: 420, loss: 0.028790703043341637
step: 430, loss: 0.010859418660402298
step: 440, loss: 0.07039058953523636
step: 450, loss: 0.16600541770458221
step: 460, loss: 0.018373999744653702
step: 470, loss: 0.10017646849155426
step: 480, loss: 0.0025000080931931734
step: 490, loss: 0.05847835913300514
step: 500, loss: 0.0059754569083452225
step: 510, loss: 0.06390537321567535
step: 520, loss: 0.012855791486799717
step: 530, loss: 0.044598985463380814
/usr/local/lib/python3.7/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=0.9269427640763147, f1=0.9184909175593852,
best_f1=0.9267618608935975
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00634364876896143
step: 10, loss: 0.0032120170071721077
step: 20, loss: 0.0011019217781722546
step: 30, loss: 0.00469936802983284
step: 40, loss: 0.024514319375157356
step: 50, loss: 0.0033908593468368053
step: 60, loss: 0.0043735140934586525
step: 70, loss: 0.011539390310645103
step: 80, loss: 0.0023641891311854124
step: 90, loss: 0.02106303721666336
step: 100, loss: 0.0008305831579491496
step: 110, loss: 0.0012009565252810717
step: 120, loss: 0.09032391011714935
step: 130, loss: 0.007205480244010687
step: 140, loss: 0.0014807234983891249
step: 150, loss: 0.0006086182547733188
step: 160, loss: 0.012806454673409462
step: 170, loss: 0.04674070328474045

```


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

```

step: 180, loss: 0.0008492148481309414
step: 190, loss: 0.004655610769987106
step: 200, loss: 0.09526145458221436
step: 210, loss: 0.0032585409935563803
step: 220, loss: 0.0013247395399957895
step: 230, loss: 0.0006665167165920138
step: 240, loss: 0.005516987759619951
step: 250, loss: 0.0018118566367775202
step: 260, loss: 0.03738817572593689
step: 270, loss: 0.022572951391339302
step: 280, loss: 0.02371811494231224
step: 290, loss: 0.03411151468753815
step: 300, loss: 0.031192364171147346
step: 310, loss: 0.009400693699717522
step: 320, loss: 0.009836261160671711
step: 330, loss: 0.009471786208450794
step: 340, loss: 0.012941480614244938
step: 350, loss: 0.05180922895669937
step: 360, loss: 0.010815640911459923
step: 370, loss: 0.0048269289545714855
step: 380, loss: 0.06213071942329407
step: 390, loss: 0.00032565079163759947
step: 400, loss: 0.007225099951028824
step: 410, loss: 0.003348414320498705
step: 420, loss: 0.029218273237347603
step: 430, loss: 0.022163938730955124
step: 440, loss: 0.0010314900428056717
step: 450, loss: 0.0013543989043682814
step: 460, loss: 0.013368869200348854
step: 470, loss: 0.004389143083244562
step: 480, loss: 0.005644817836582661
step: 490, loss: 0.003242185339331627
step: 500, loss: 0.03500143066048622
step: 510, loss: 0.003147359471768141
step: 520, loss: 0.008398081175982952
step: 530, loss: 0.017994381487369537
/usr/local/lib/python3.7/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=0.9268978444236176, f1=0.9315960912052118,
best_f1=0.9267618608935975
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04074176400899887
step: 10, loss: 0.005819974932819605
step: 20, loss: 0.001662357128225267

```

step: 30, loss: 0.00614518066868186
step: 40, loss: 0.002538933651521802
step: 50, loss: 0.0023264847695827484
step: 60, loss: 0.002971308771520853
step: 70, loss: 0.02911338396370411
step: 80, loss: 0.0005002759280614555
step: 90, loss: 0.0010653305798768997
step: 100, loss: 0.0021543465554714203
step: 110, loss: 0.006776876747608185
step: 120, loss: 0.004113560542464256
step: 130, loss: 0.0009668441489338875
step: 140, loss: 0.058839134871959686
step: 150, loss: 0.0007454018923453987
step: 160, loss: 0.0018931925296783447
step: 170, loss: 0.0005810578004457057
step: 180, loss: 0.001423163223080337
step: 190, loss: 0.01354297623038292
step: 200, loss: 0.00042506400495767593
step: 210, loss: 0.008165119215846062
step: 220, loss: 0.00102412945125252
step: 230, loss: 0.0025461274199187756
step: 240, loss: 0.0005453019984997809
step: 250, loss: 0.00014626613119617105
step: 260, loss: 0.00045634369598701596
step: 270, loss: 0.0002096703101415187
step: 280, loss: 0.00510892691090703
step: 290, loss: 0.014410365372896194
step: 300, loss: 0.002317231148481369
step: 310, loss: 0.002372661605477333
step: 320, loss: 0.002469351515173912
step: 330, loss: 0.0003173378063365817
step: 340, loss: 0.029193438589572906
step: 350, loss: 0.02414911426603794
step: 360, loss: 0.0003459393628872931
step: 370, loss: 0.030524887144565582
step: 380, loss: 0.0024346557911485434
step: 390, loss: 0.0488666296005249
step: 400, loss: 0.0008067050366662443
step: 410, loss: 0.035819023847579956
step: 420, loss: 0.030594421550631523
step: 430, loss: 0.003582785138860345
step: 440, loss: 0.020314905792474747
step: 450, loss: 0.035848986357450485
step: 460, loss: 0.007617482915520668
step: 470, loss: 0.08694222569465637
step: 480, loss: 0.0035887067206203938
step: 490, loss: 0.0003245931293349713
step: 500, loss: 0.009213504381477833

```
step: 510, loss: 0.00011965398880420253
step: 520, loss: 0.001980534987524152
step: 530, loss: 0.014328928664326668
/usr/local/lib/python3.7/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=0.9198672356567094, f1=0.9119318181818182,
best_f1=0.9267618608935975
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00626123882830143
step: 10, loss: 0.010488461703062057
step: 20, loss: 0.005532821640372276
step: 30, loss: 0.002432743553072214
step: 40, loss: 0.006040408741682768
step: 50, loss: 0.0005142265581525862
step: 60, loss: 0.06247284263372421
step: 70, loss: 0.01543776411563158
step: 80, loss: 0.004486734978854656
step: 90, loss: 0.0007152976468205452
step: 100, loss: 0.0006884750328026712
step: 110, loss: 0.0053099519573152065
step: 120, loss: 0.00012085203343303874
step: 130, loss: 0.0031868633814156055
step: 140, loss: 0.0010085197864100337
step: 150, loss: 0.007640352938324213
step: 160, loss: 0.0008775845635682344
step: 170, loss: 0.004626549780368805
step: 180, loss: 0.004097884055227041
step: 190, loss: 0.0009805381996557117
step: 200, loss: 0.028855521231889725
step: 210, loss: 0.003963345196098089
step: 220, loss: 0.00019051085109822452
step: 230, loss: 0.00014796172035858035
step: 240, loss: 0.0013395952992141247
step: 250, loss: 0.015999644994735718
step: 260, loss: 0.0020347791723906994
step: 270, loss: 0.009659103117883205
step: 280, loss: 0.004069655202329159
step: 290, loss: 0.002531952690333128
step: 300, loss: 0.00013561589003074914
step: 310, loss: 0.000654625880997628
step: 320, loss: 0.06420019268989563
step: 330, loss: 0.0042930422350764275
step: 340, loss: 0.03465033322572708
step: 350, loss: 0.001770773553289473
step: 360, loss: 0.010624650865793228
```

```

step: 370, loss: 0.0019455800065770745
step: 380, loss: 0.03309870883822441
step: 390, loss: 0.00010419765749247745
step: 400, loss: 0.0019472521962597966
step: 410, loss: 0.0010544859105721116
step: 420, loss: 0.0001837550662457943
step: 430, loss: 0.0002241588372271508
step: 440, loss: 0.00022834824630990624
step: 450, loss: 0.024494759738445282
step: 460, loss: 0.0025960092898458242
step: 470, loss: 0.07196182012557983
step: 480, loss: 0.003922739531844854
step: 490, loss: 0.007620367221534252
step: 500, loss: 0.0006948431837372482
step: 510, loss: 0.0017653449904173613
step: 520, loss: 0.009931440465152264
step: 530, loss: 0.00036112911766394973
/usr/local/lib/python3.7/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=0.9278538812785389, f1=0.9196141479099679,
best_f1=0.9267618608935975
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.006184547673910856
step: 10, loss: 0.07122842967510223
step: 20, loss: 0.0228725615888834
step: 30, loss: 0.008020631037652493
step: 40, loss: 0.0009599464829079807
step: 50, loss: 0.000416740047512576
step: 60, loss: 4.812519910046831e-05
step: 70, loss: 0.0009362206910736859
step: 80, loss: 0.0001429479889338836
step: 90, loss: 0.07575348764657974
step: 100, loss: 0.008691255003213882
step: 110, loss: 0.0032020285725593567
step: 120, loss: 0.002125814789906144
step: 130, loss: 0.002197525231167674
step: 140, loss: 4.156530849286355e-05
step: 150, loss: 0.0030280796345323324
step: 160, loss: 0.0025558883789926767
step: 170, loss: 0.06173868477344513
step: 180, loss: 0.000710715539753437
step: 190, loss: 0.00030520273139700294
step: 200, loss: 0.0002843473630491644
step: 210, loss: 0.0041274684481322765
step: 220, loss: 0.008056042715907097

```

```

step: 230, loss: 0.0031596114858984947
step: 240, loss: 8.497077942593023e-05
step: 250, loss: 0.004016416147351265
step: 260, loss: 0.01212482899427414
step: 270, loss: 8.649556548334658e-05
step: 280, loss: 9.644454257795587e-05
step: 290, loss: 0.004174342844635248
step: 300, loss: 0.00014321938215289265
step: 310, loss: 0.018549617379903793
step: 320, loss: 0.0003104110073763877
step: 330, loss: 0.020286237820982933
step: 340, loss: 0.03414520248770714
step: 350, loss: 0.00030466122552752495
step: 360, loss: 0.0004438321047928184
step: 370, loss: 0.018168997019529343
step: 380, loss: 0.000497666303999722
step: 390, loss: 0.0003790612972807139
step: 400, loss: 0.1426822692155838
step: 410, loss: 0.0003492622636258602
step: 420, loss: 0.00010290119826095179
step: 430, loss: 0.00026770285330712795
step: 440, loss: 0.011693638749420643
step: 450, loss: 0.0007379024755209684
step: 460, loss: 0.0050678919069468975
step: 470, loss: 0.004420200362801552
step: 480, loss: 0.00026062544202432036
step: 490, loss: 0.0001768084184732288
step: 500, loss: 0.003617049427703023
step: 510, loss: 0.00018349243327975273
step: 520, loss: 0.00017931347247213125
step: 530, loss: 0.00034946456435136497
/usr/local/lib/python3.7/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=0.9264029438822446, f1=0.9255514705882354,
best_f1=0.9267618608935975
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0008275876753032207
step: 10, loss: 0.1421404480934143
step: 20, loss: 0.0007860997575335205
step: 30, loss: 0.004518060013651848
step: 40, loss: 0.0075794183649122715
step: 50, loss: 0.0002614678523968905
step: 60, loss: 0.0019746068865060806
step: 70, loss: 0.12680034339427948
step: 80, loss: 0.0005263070925138891

```

```

step: 90, loss: 0.0272025465965271
step: 100, loss: 0.0015406584134325385
step: 110, loss: 0.02342730388045311
step: 120, loss: 0.02771322801709175
step: 130, loss: 0.00010016022861236706
step: 140, loss: 0.0009214184246957302
step: 150, loss: 0.035807881504297256
step: 160, loss: 0.001202564686536789
step: 170, loss: 0.0003820787533186376
step: 180, loss: 4.0428571082884446e-05
step: 190, loss: 0.004829375073313713
step: 200, loss: 0.003262563608586788
step: 210, loss: 0.0001859050098573789
step: 220, loss: 0.0014686171198263764
step: 230, loss: 0.0003438517451286316
step: 240, loss: 0.00016354113176930696
step: 250, loss: 0.013438915833830833
step: 260, loss: 6.32485025562346e-05
Gradient overflow. Skipping step, loss scaler 0 reducing loss scale to 16384.0
step: 270, loss: 0.0006940409075468779
step: 280, loss: 3.391402424313128e-05
step: 290, loss: 4.502825322560966e-05
step: 300, loss: 0.009123019874095917
step: 310, loss: 8.173599053407088e-05
step: 320, loss: 8.043283014558256e-05
step: 330, loss: 0.0011665495112538338
step: 340, loss: 6.395943637471646e-05
step: 350, loss: 0.0002597508719190955
step: 360, loss: 0.022024981677532196
step: 370, loss: 0.00032870614086277783
step: 380, loss: 4.312838791520335e-05
step: 390, loss: 0.0006679992075078189
step: 400, loss: 0.0022832658141851425
step: 410, loss: 0.00048172875540331006
step: 420, loss: 0.0006567409145645797
step: 430, loss: 7.23715202184394e-05
step: 440, loss: 8.363308006664738e-05
step: 450, loss: 0.00010040056804427877
step: 460, loss: 0.034346986562013626
step: 470, loss: 0.0009546208893880248
step: 480, loss: 0.009549098089337349
step: 490, loss: 0.0002959470439236611
step: 500, loss: 0.12450267374515533
step: 510, loss: 0.0005667575169354677
step: 520, loss: 0.000265899405349046
step: 530, loss: 0.0019095811294391751
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=0.927816091954023, f1=0.9286700414173953,
best_f1=0.9267618608935975
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.04784044623374939
step: 10, loss: 0.00041197650716640055
step: 20, loss: 0.0005149816861376166
step: 30, loss: 0.0005122960428707302
step: 40, loss: 0.0015649686101824045
step: 50, loss: 0.0001799566380213946
step: 60, loss: 0.000519879802595824
step: 70, loss: 0.0005069179460406303
step: 80, loss: 0.00148701888974756
step: 90, loss: 0.0008034948259592056
step: 100, loss: 0.0002790311991702765
step: 110, loss: 0.0016492882277816534
step: 120, loss: 0.0006666444824077189
step: 130, loss: 0.0030178287997841835
step: 140, loss: 0.06265649944543839
step: 150, loss: 0.00014051669859327376
step: 160, loss: 0.0004481151991058141
step: 170, loss: 0.000318740785587579
step: 180, loss: 0.11381321400403976
step: 190, loss: 0.00030794087797403336
step: 200, loss: 0.0003402724105399102
step: 210, loss: 5.117570617585443e-05
step: 220, loss: 0.00032960809767246246
step: 230, loss: 0.000758378766477108
step: 240, loss: 0.00017951651534531265
step: 250, loss: 0.00035456879413686693
step: 260, loss: 0.006256898399442434
step: 270, loss: 0.00017104212020058185
step: 280, loss: 8.266355871455744e-05
step: 290, loss: 0.0002907172020059079
step: 300, loss: 0.06271573156118393
step: 310, loss: 0.0001754390395944938
step: 320, loss: 0.00015342063852585852
step: 330, loss: 0.0002469090395607054
step: 340, loss: 0.0016568801365792751
step: 350, loss: 4.084920510649681e-05
step: 360, loss: 0.0010770033113658428
step: 370, loss: 0.0006136354058980942
step: 380, loss: 0.00010525633115321398
step: 390, loss: 4.1459701606072485e-05
step: 400, loss: 0.0001607544836588204

```

```

step: 410, loss: 0.0010118078207597136
step: 420, loss: 0.00040946516674011946
step: 430, loss: 0.0009277951903641224
step: 440, loss: 8.510831685271114e-05
step: 450, loss: 0.0019025482470169663
step: 460, loss: 0.00830040592700243
step: 470, loss: 0.00023930612951517105
step: 480, loss: 0.005732231307774782
step: 490, loss: 0.04117352142930031
step: 500, loss: 0.05969243496656418
step: 510, loss: 0.0005281081539578736
step: 520, loss: 0.00043971356353722513
step: 530, loss: 0.007063302211463451
/usr/local/lib/python3.7/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=0.9311141932501156, f1=0.9253592953175707,
best_f1=0.9267618608935975
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0014710661489516497
step: 10, loss: 0.008471701294183731
step: 20, loss: 0.0012412493815645576
step: 30, loss: 0.004416916519403458
step: 40, loss: 0.00043945389916189015
step: 50, loss: 0.0010566223645582795
step: 60, loss: 8.625844202470034e-05
step: 70, loss: 0.00012579845497384667
step: 80, loss: 0.00019962973601650447
step: 90, loss: 0.0008957530953921378
step: 100, loss: 5.611731830867939e-05
step: 110, loss: 5.604462785413489e-05
step: 120, loss: 0.0033804054837673903
step: 130, loss: 0.00021521247981581837
step: 140, loss: 7.630565960425884e-05
step: 150, loss: 8.106623135972768e-05
step: 160, loss: 0.048793382942676544
step: 170, loss: 0.007690641563385725
step: 180, loss: 0.00012301633250899613
step: 190, loss: 0.01687162183225155
step: 200, loss: 0.0010137358913198113
step: 210, loss: 0.000539142987690866
step: 220, loss: 0.0006499215378426015
step: 230, loss: 0.0012612747959792614
step: 240, loss: 0.0005350019782781601
step: 250, loss: 0.0009940232848748565
step: 260, loss: 0.000231096419156529

```



```

step: 270, loss: 0.01726723089814186
step: 280, loss: 0.001206427114084363
step: 290, loss: 0.009333171881735325
step: 300, loss: 0.05910877883434296
step: 310, loss: 0.013614978641271591
step: 320, loss: 0.0260151457041502
step: 330, loss: 0.0022306623868644238
step: 340, loss: 0.00016755113028921187
step: 350, loss: 0.0018650952260941267
step: 360, loss: 0.0006234440952539444
step: 370, loss: 0.00032045593252405524
step: 380, loss: 0.00015351854381151497
step: 390, loss: 0.0007240294362418354
step: 400, loss: 6.69631190248765e-05
step: 410, loss: 0.0013326086336746812
step: 420, loss: 0.016212932765483856
step: 430, loss: 0.0005952564533799887
step: 440, loss: 0.00021674252639058977
step: 450, loss: 0.00012473046081140637
step: 460, loss: 0.0010942477965727448
step: 470, loss: 0.0005115381209179759
step: 480, loss: 6.171203858684748e-05
step: 490, loss: 0.0726698637008667
step: 500, loss: 0.0001518219942227006
step: 510, loss: 7.746028859401122e-05
step: 520, loss: 0.0011150463251397014
step: 530, loss: 0.005991601385176182
/usr/local/lib/python3.7/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=0.9313815187557181, f1=0.9303391384051329,
best_f1=0.9267618608935975
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
  warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.00033315346809104085
step: 10, loss: 0.0010625893482938409
step: 20, loss: 0.0007612240733578801
step: 30, loss: 0.06122135370969772
step: 40, loss: 0.00010603573900880292
step: 50, loss: 0.0025452503468841314
step: 60, loss: 0.0002295735466759652
step: 70, loss: 0.04326947033405304
step: 80, loss: 0.0013778763823211193
step: 90, loss: 0.0005628567305393517
step: 100, loss: 0.006483669392764568
step: 110, loss: 0.0011253055417910218
step: 120, loss: 0.0005344273522496223

```

```

step: 130, loss: 0.0003449738724157214
step: 140, loss: 0.00018649805861059576
step: 150, loss: 0.000478848785860464
step: 160, loss: 0.0002454481436870992
step: 170, loss: 0.00014760061458218843
step: 180, loss: 4.472062937566079e-05
step: 190, loss: 3.729318632395007e-05
step: 200, loss: 0.00031472250702790916
step: 210, loss: 0.0005968717159703374
step: 220, loss: 2.678761120478157e-05
step: 230, loss: 0.0006290010642260313
step: 240, loss: 0.00025234438362531364
step: 250, loss: 3.8125075661810115e-05
step: 260, loss: 0.007143632974475622
step: 270, loss: 0.005161784123629332
step: 280, loss: 3.2885251130210236e-05
step: 290, loss: 0.00283572799526155
step: 300, loss: 0.0005063794087618589
step: 310, loss: 6.370037590386346e-05
step: 320, loss: 0.0030676841270178556
step: 330, loss: 6.555977597599849e-05
step: 340, loss: 0.0009781334083527327
step: 350, loss: 0.02479930780827999
step: 360, loss: 0.009458792395889759
step: 370, loss: 0.0004327698261477053
step: 380, loss: 5.397434506448917e-05
step: 390, loss: 2.4362878320971504e-05
step: 400, loss: 0.029678193852305412
step: 410, loss: 0.0007325204787775874
step: 420, loss: 0.00031520056654699147
step: 430, loss: 0.00016247824532911181
step: 440, loss: 0.0015523217152804136
step: 450, loss: 0.005128929857164621
step: 460, loss: 0.0004697303520515561
step: 470, loss: 0.00016256920935120434
step: 480, loss: 0.0002700370969250798
step: 490, loss: 0.005788112990558147
step: 500, loss: 0.0015715904301032424
step: 510, loss: 0.03740601986646652
step: 520, loss: 0.001964869210496545
step: 530, loss: 0.0008295496227219701
/usr/local/lib/python3.7/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=0.9297197978870005, f1=0.9255514705882354,
best_f1=0.9267618608935975
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.

```

```
warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0028233963530510664
step: 10, loss: 0.0004831255355384201
step: 20, loss: 0.0038767268415540457
step: 30, loss: 0.00011162170994793996
step: 40, loss: 5.508919639396481e-05
step: 50, loss: 7.39126480766572e-05
step: 60, loss: 6.807954923715442e-05
step: 70, loss: 8.163589518517256e-05
step: 80, loss: 4.602935223374516e-05
step: 90, loss: 0.00015783667913638055
step: 100, loss: 0.00021038226259406656
step: 110, loss: 0.0043347240425646305
step: 120, loss: 9.091373067349195e-05
step: 130, loss: 0.00039289952837862074
step: 140, loss: 0.001514787902124226
step: 150, loss: 0.0001762316096574068
step: 160, loss: 0.0008572295191697776
step: 170, loss: 0.0002771921281237155
step: 180, loss: 0.000869304989464581
step: 190, loss: 0.0028767248149961233
step: 200, loss: 0.003310183063149452
step: 210, loss: 0.0016633053310215473
step: 220, loss: 0.00013541466614697129
step: 230, loss: 0.0027148155495524406
step: 240, loss: 7.958858623169363e-05
step: 250, loss: 0.022834422066807747
step: 260, loss: 0.05554094538092613
step: 270, loss: 0.011353982612490654
step: 280, loss: 0.0035113247577100992
step: 290, loss: 0.0010035685263574123
step: 300, loss: 0.0011565107852220535
step: 310, loss: 0.000474553118692711
step: 320, loss: 0.0005392725579440594
step: 330, loss: 0.011743085458874702
step: 340, loss: 5.643114127451554e-05
step: 350, loss: 0.0021174608264118433
step: 360, loss: 6.631162978010252e-05
step: 370, loss: 0.00024491449585184455
step: 380, loss: 0.0010461148340255022
step: 390, loss: 0.00022141473891679198
step: 400, loss: 0.00022081514180172235
step: 410, loss: 0.00012329213495831937
step: 420, loss: 0.0002379057987127453
step: 430, loss: 0.000407327723223716
step: 440, loss: 7.847156666684896e-05
step: 450, loss: 0.10254930704832077
step: 460, loss: 2.3099895770428702e-05
```

```

step: 470, loss: 0.01957794837653637
step: 480, loss: 0.0013995547778904438
step: 490, loss: 0.0038177266251295805
step: 500, loss: 0.0020505087450146675
step: 510, loss: 4.9297846999252215e-05
step: 520, loss: 0.00031142824445851147
step: 530, loss: 7.174171332735568e-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=0.9297912713472486, f1=0.9217638691322902,
best_f1=0.9267618608935975
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 0.0006968635716475546
step: 10, loss: 0.0008640268933959305
step: 20, loss: 3.5384251532377675e-05
step: 30, loss: 0.000791239261161536
step: 40, loss: 0.0021246266551315784
step: 50, loss: 0.00017987073806580156
step: 60, loss: 8.1061982200481e-05
step: 70, loss: 3.1290041079046205e-05
step: 80, loss: 0.008394671604037285
step: 90, loss: 8.331233402714133e-05
step: 100, loss: 1.7203039533342235e-05
step: 110, loss: 4.4933116441825405e-05
step: 120, loss: 5.366188270272687e-05
step: 130, loss: 5.414645420387387e-05
step: 140, loss: 0.006404577754437923
step: 150, loss: 0.06916149705648422
step: 160, loss: 0.0016209500608965755
step: 170, loss: 0.00028767395997419953
step: 180, loss: 0.0004627889720723033
step: 190, loss: 0.00032609220943413675
step: 200, loss: 0.0009148052195087075
step: 210, loss: 0.006411731243133545
step: 220, loss: 0.0015435813693329692
step: 230, loss: 0.0014776577008888125
step: 240, loss: 0.08381588011980057
step: 250, loss: 0.003441826906055212
step: 260, loss: 4.1750998207135126e-05
step: 270, loss: 0.0019162304233759642
step: 280, loss: 0.0037377893459051847
step: 290, loss: 8.010832971194759e-05
step: 300, loss: 0.00046299188397824764
step: 310, loss: 0.0007919264025986195
step: 320, loss: 0.0005654317792505026

```

```
step: 330, loss: 0.0014187927590683103
step: 340, loss: 0.00018828091560862958
step: 350, loss: 0.0019234042847529054
step: 360, loss: 0.0017411697190254927
step: 370, loss: 0.00048599657020531595
step: 380, loss: 0.00010054075391963124
step: 390, loss: 0.00035171560011804104
step: 400, loss: 2.09169684239896e-05
step: 410, loss: 8.370848081540316e-05
step: 420, loss: 0.0003598856565076858
step: 430, loss: 0.0002523158327676356
step: 440, loss: 2.7398395104683004e-05
step: 450, loss: 0.0008980737184174359
step: 460, loss: 0.23790676891803741
step: 470, loss: 4.836400330532342e-05
step: 480, loss: 2.8639236916205846e-05
step: 490, loss: 0.00015946671192068607
step: 500, loss: 0.00012805970618501306
step: 510, loss: 0.022884100675582886
step: 520, loss: 0.00020883367687929422
step: 530, loss: 0.0001004710138658993
/usr/local/lib/python3.7/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=0.928639391056137, f1=0.9178605539637059,
best_f1=0.9267618608935975
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
step: 0, loss: 7.32989574316889e-05
step: 10, loss: 7.898537296568975e-05
step: 20, loss: 0.0007462709327228367
step: 30, loss: 0.00036006077425554395
step: 40, loss: 3.0102744858595543e-05
step: 50, loss: 6.93281544954516e-05
step: 60, loss: 0.0002957867691293359
step: 70, loss: 0.00016456506273243576
step: 80, loss: 6.331294571282342e-05
step: 90, loss: 2.6085670469910838e-05
step: 100, loss: 2.3416610929416493e-05
step: 110, loss: 0.025404417887330055
step: 120, loss: 0.0014481327962130308
step: 130, loss: 0.00016368782962672412
step: 140, loss: 1.8037417248706333e-05
step: 150, loss: 0.0014194012619554996
step: 160, loss: 6.242527888389304e-05
step: 170, loss: 0.00015180677291937172
step: 180, loss: 6.549269164679572e-05
```

```

step: 190, loss: 0.00019806748605333269
step: 200, loss: 0.009700017981231213
step: 210, loss: 7.608626765431836e-05
step: 220, loss: 4.765187986777164e-05
step: 230, loss: 0.00020938234229106456
step: 240, loss: 7.784656918374822e-05
step: 250, loss: 0.0003500230668578297
step: 260, loss: 0.0003094166750088334
step: 270, loss: 0.0009862416191026568
step: 280, loss: 0.001043380587361753
step: 290, loss: 0.0013435856672003865
step: 300, loss: 0.0008768816478550434
step: 310, loss: 0.0009473515674471855
step: 320, loss: 0.0003486968926154077
step: 330, loss: 8.320584311150014e-05
step: 340, loss: 8.902591798687354e-05
step: 350, loss: 0.00044319790322333574
step: 360, loss: 0.00024984890478663146
step: 370, loss: 2.2023226847522892e-05
step: 380, loss: 0.00014324250514619052
step: 390, loss: 0.00022644399723503739
step: 400, loss: 0.00026437139604240656
step: 410, loss: 0.00013636417861562222
step: 420, loss: 8.358601189684123e-05
step: 430, loss: 0.00010100970393978059
step: 440, loss: 0.007921511307358742
step: 450, loss: 8.979648555396125e-05
step: 460, loss: 9.976040746551007e-05
step: 470, loss: 7.944431126816198e-05
step: 480, loss: 0.00023262877948582172
step: 490, loss: 0.0004910138086415827
step: 500, loss: 0.0006330955657176673
step: 510, loss: 8.150090434355661e-05
step: 520, loss: 0.0037085937801748514
step: 530, loss: 2.2537516997545026e-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=0.9288702928870293, f1=0.9288040949278734,
best_f1=0.9267618608935975
mv: cannot stat '*_dev.pt': No such file or directory
Selected optimization level 02:  FP16 training with FP32 batchnorm and FP32
master weights.

Defaults for this optimization level are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16

```

```

patch_torch_functions : False
keep_batchnorm_fp32   : True
master_weights        : True
loss_scale             : dynamic
Processing user overrides (additional kwargs that are not None)...
After processing overrides, optimization options are:
enabled                : True
opt_level              : 02
cast_model_type        : torch.float16
patch_torch_functions : False
keep_batchnorm_fp32   : True
master_weights        : True
loss_scale             : dynamic
Warning: multi_tensor_applier fused unscale kernel is unavailable, possibly
because apex was installed without --cuda_ext --cpp_ext. Using Python fallback.
Original ImportError was: ModuleNotFoundError("No module named 'amp_C'")
/usr/local/lib/python3.7/dist-packages/apex/amp/_initialize.py:25: UserWarning:
An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
5742it [00:14, 405.33it/s]
load_f1 = 0.9323515876668201
real_f1 = 0.9321100917431193
Oit [00:00, ?it/s]/usr/local/lib/python3.7/dist-
packages/apex/amp/_initialize.py:25: UserWarning: An input tensor was not cuda.
    warnings.warn("An input tensor was not cuda.")
4398it [00:10, 414.04it/s]

```