

HW4

November 11, 2023

1 Task 1

1.0.1 What are the precision, recall, and F1 score on the validation data?

Precision: 82.36% Recall: 76.39% F1 score: 79.04 on the best run

1.0.2 What are the precision, recall, and F1 score on the test data?

Precision: 74.25% Recall: 68.25% F1 score: 71.13 with the model from the best run

2 Task 2

2.0.1 What is the precision, recall, and F1 score on the validation data?

Precision: 89.27% Recall: 89.48% F1 score: 89.37 on the best run

2.0.2 What are the precision, recall, and F1 score on the test data?

Precision: 84.33% Recall: 85.09% F1 score: 84.71 with the model from the best run

2.0.3 BiLSTM with GloVe Embeddings outperforms the model without. Can you provide a rationale for this?

There are two main reasons for this observation. - The GloVe embeddings are trained on a corpora of around 400,000 words compared to the 23,623 words from the training dataset. This means the GloVe model will understand a lot more words and replace a lot fewer with an unknown tag. - The GloVe embeddings also encode semantic information. They embeddings will have contextual information and will therefore, in a sense, be able to understand the meaning and context of words.

```
[ ]: !pip install datasets accelerate
      !wget https://raw.githubusercontent.com/sighsmile/conlleva1/master/conlleva1.py
```

Collecting datasets

Downloading datasets-2.14.6-py3-none-any.whl (493 kB)
493.7/493.7

kB 8.3 MB/s eta 0:00:00

Collecting accelerate

Downloading accelerate-0.24.1-py3-none-any.whl (261 kB)

261.4/261.4

kB 12.7 MB/s eta 0:00:00

Requirement already satisfied: numpy>=1.17 in

/usr/local/lib/python3.10/dist-packages (from datasets) (1.23.5)

Requirement already satisfied: pyarrow>=8.0.0 in /usr/local/lib/python3.10/dist-packages (from datasets) (9.0.0)

Collecting dill<0.3.8,>=0.3.0 (from datasets)

Downloading dill-0.3.7-py3-none-any.whl (115 kB)

115.3/115.3

kB 8.2 MB/s eta 0:00:00

Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from datasets) (1.5.3)

Requirement already satisfied: requests>=2.19.0 in

/usr/local/lib/python3.10/dist-packages (from datasets) (2.31.0)

Requirement already satisfied: tqdm>=4.62.1 in /usr/local/lib/python3.10/dist-packages (from datasets) (4.66.1)

Requirement already satisfied: xxhash in /usr/local/lib/python3.10/dist-packages (from datasets) (3.4.1)

Collecting multiprocessing (from datasets)

Downloading multiprocessing-0.70.15-py310-none-any.whl (134 kB)

134.8/134.8

kB 11.5 MB/s eta 0:00:00

Requirement already satisfied: fsspec[http]<=2023.10.0,>=2023.1.0 in

/usr/local/lib/python3.10/dist-packages (from datasets) (2023.6.0)

Requirement already satisfied: aiohttp in /usr/local/lib/python3.10/dist-packages (from datasets) (3.8.6)

Collecting huggingface-hub<1.0.0,>=0.14.0 (from datasets)

Downloading huggingface_hub-0.19.0-py3-none-any.whl (311 kB)

311.2/311.2

kB 15.5 MB/s eta 0:00:00

Requirement already satisfied: packaging in

/usr/local/lib/python3.10/dist-packages (from datasets) (23.2)

Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from datasets) (6.0.1)

Requirement already satisfied: psutil in /usr/local/lib/python3.10/dist-packages (from accelerate) (5.9.5)

Requirement already satisfied: torch>=1.10.0 in /usr/local/lib/python3.10/dist-packages (from accelerate) (2.1.0+cu118)

Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (23.1.0)

Requirement already satisfied: charset-normalizer<4.0,>=2.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (3.3.2)

Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (6.0.4)

Requirement already satisfied: async-timeout<5.0,>=4.0.0a3 in

```

/usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (4.0.3)
Requirement already satisfied: yarl<2.0,>=1.0 in /usr/local/lib/python3.10/dist-
packages (from aiohttp->datasets) (1.9.2)
Requirement already satisfied: frozenlist>=1.1.1 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (1.4.0)
Requirement already satisfied: aiosignal>=1.1.2 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (1.3.1)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-
packages (from huggingface-hub<1.0.0,>=0.14.0->datasets) (3.13.1)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/usr/local/lib/python3.10/dist-packages (from huggingface-
hub<1.0.0,>=0.14.0->datasets) (4.5.0)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-
packages (from requests>=2.19.0->datasets) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.19.0->datasets)
(2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.19.0->datasets)
(2023.7.22)
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages
(from torch>=1.10.0->accelerate) (1.12)
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-
packages (from torch>=1.10.0->accelerate) (3.2.1)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages
(from torch>=1.10.0->accelerate) (3.1.2)
Requirement already satisfied: triton==2.1.0 in /usr/local/lib/python3.10/dist-
packages (from torch>=1.10.0->accelerate) (2.1.0)
Requirement already satisfied: python-dateutil>=2.8.1 in
/usr/local/lib/python3.10/dist-packages (from pandas->datasets) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-
packages (from pandas->datasets) (2023.3.post1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-
packages (from python-dateutil>=2.8.1->pandas->datasets) (1.16.0)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.10/dist-packages (from jinja2->torch>=1.10.0->accelerate)
(2.1.3)
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/dist-
packages (from sympy->torch>=1.10.0->accelerate) (1.3.0)
Installing collected packages: dill, multiprocessing, huggingface-hub, accelerate,
datasets
Successfully installed accelerate-0.24.1 datasets-2.14.6 dill-0.3.7 huggingface-
hub-0.19.0 multiprocessing-0.70.15
--2023-11-11 00:53:18--
https://raw.githubusercontent.com/sighsmlr/conlleval/master/conlleval.py
Resolving raw.githubusercontent.com (raw.githubusercontent.com)...
185.199.108.133, 185.199.109.133, 185.199.110.133, ...
Connecting to raw.githubusercontent.com

```

(raw.githubusercontent.com)|185.199.108.133|:443... connected.

HTTP request sent, awaiting response... 200 OK

Length: 7502 (7.3K) [text/plain]

Saving to: 'conlleval.py'

conlleval.py 100%[=====>] 7.33K --.-KB/s in 0s

2023-11-11 00:53:18 (81.8 MB/s) - 'conlleval.py' saved [7502/7502]

```
[ ]: import datasets
import numpy as np
import torch
import torch.nn as nn
import copy
from torch.utils.data import DataLoader
from accelerate import Accelerator
from torch.optim.lr_scheduler import ReduceLROnPlateau
import itertools
from collections import Counter
from conlleval import evaluate
```

```
[ ]: DATASET_NAME = 'conll2003'
UNKNOWN = '[UNK]'
PAD = '[PAD]'
PAD_TOKEN_IDX = 0
```

```
[ ]: tag2idx = {'O': 0, 'B-PER': 1, 'I-PER': 2, 'B-ORG': 3, 'I-ORG': 4, 'B-LOC': 5,
↳ 'I-LOC': 6, 'B-MISC': 7, 'I-MISC': 8}
tags = list(tag2idx.keys())
tag2idx[PAD] = 9
```

```
[ ]: dataset = datasets.load_dataset(DATASET_NAME)
```

Downloading builder script: 0%| | 0.00/9.57k [00:00<?, ?B/s]

Downloading metadata: 0%| | 0.00/3.73k [00:00<?, ?B/s]

Downloading readme: 0%| | 0.00/12.3k [00:00<?, ?B/s]

Downloading data: 0%| | 0.00/983k [00:00<?, ?B/s]

Generating train split: 0%| | 0/14041 [00:00<?, ? examples/s]

Generating validation split: 0%| | 0/3250 [00:00<?, ? examples/s]

Generating test split: 0%| | 0/3453 [00:00<?, ? examples/s]

2.1 Task 1 training

```
[ ]: word2idx = Counter(itertools.chain(*dataset['train']['tokens']))

word2idx = {
    word: frequency
    for word, frequency in word2idx.items()
    if frequency >= 3
}

word2idx = {
    word: index
    for index, word in enumerate(word2idx.keys(), start=2)
}

word2idx[PAD] = PAD_TOKEN_IDX
word2idx[UNKNOWN] = 1

[ ]: def preprocess_sample_task1(sample):
    sample['input_ids'] = [word2idx.get(token, word2idx[UNKNOWN]) for token in
        ↪sample['tokens']]

    del sample['pos_tags']
    del sample['chunk_tags']
    del sample['tokens']
    del sample['id']

    sample['labels'] = sample['ner_tags']
    del sample['ner_tags']

    return sample

dataset_task1 = dataset.map(preprocess_sample_task1)

Map:   0%|          | 0/14041 [00:00<?, ? examples/s]
Map:   0%|          | 0/3250 [00:00<?, ? examples/s]
Map:   0%|          | 0/3453 [00:00<?, ? examples/s]

[ ]: INPUT_DIMENSION = len(word2idx)
    EMBEDDING_DIMENSION = 100
    NUM_LSTM_LAYER = 1
    LSTM_HIDDEN_DIMENSION = 256
    LSTM_DROPOUT = 0.33
    LINEAR_OUTPUT_DIMENSION = 128
    NUM_LABELS = 9
```

```
[ ]: def collate_fn(batch):
    input_ids = [torch.LongTensor(sample['input_ids']) for sample in batch]
    labels = [torch.LongTensor(sample['labels']) for sample in batch]
    input_ids = nn.utils.rnn.pad_sequence(
        input_ids, batch_first=True, padding_value=PAD_TOKEN_IDX)
    labels = nn.utils.rnn.pad_sequence(
        labels, batch_first=True, padding_value=tag2idx[PAD])
    # Store the original lengths of sequences so we can get the original length
    ↪back for prediction evaluation.
    input_lengths = [len(sample['input_ids']) for sample in batch]
    return {
        'input_ids': input_ids,
        'labels': labels,
        'input_lengths': input_lengths,
    }
```

```
[ ]: batch_size = 32
train_dataset = dataset_task1['train']
val_dataset = dataset_task1['validation']

train_loader = DataLoader(train_dataset, batch_size=batch_size, shuffle=True,
    ↪collate_fn=collate_fn)
val_loader = DataLoader(val_dataset, batch_size=batch_size,
    ↪collate_fn=collate_fn)
```

```
[ ]: class BiLSTMModel(nn.Module):
    def __init__(self, embedding_dim, embedding_layer, lstm_hidden_dim,
    ↪lstm_num_layers, lstm_dropout, linear_output_dim, num_labels):
        super(BiLSTMModel, self).__init__()

        self.embedding = embedding_layer
        self.bilstm = nn.LSTM(embedding_dim, lstm_hidden_dim,
    ↪num_layers=lstm_num_layers, batch_first=True, bidirectional=True)
        self.dropout = nn.Dropout(lstm_dropout)
        self.linear = nn.Linear(lstm_hidden_dim * 2, linear_output_dim)
        self.elu = nn.ELU()
        self.classifier = nn.Linear(linear_output_dim, num_labels)

    def forward(self, input_ids):
        embeddings = self.embedding(input_ids)
        lstm_output, _ = self.bilstm(embeddings)
        lstm_output = self.dropout(lstm_output)
        linear_output = self.linear(lstm_output)
        elu_output = self.elu(linear_output)
        logits = self.classifier(elu_output)
        return logits
```

```
[ ]: learning_rate = 0.001
embedding_layer = nn.Embedding(INPUT_DIMENSION, EMBEDDING_DIMENSION)
model = BiLSTMModel(EMBEDDING_DIMENSION, embedding_layer,
    ↳LSTM_HIDDEN_DIMENSION, NUM_LSTM_LAYER, LSTM_DROPOUT, LINEAR_OUTPUT_DIMENSION,
    ↳NUM_LABELS)
criterion = nn.CrossEntropyLoss(ignore_index=tag2idx[PAD])
optimizer = torch.optim.AdamW(model.parameters(), lr=learning_rate)
```

```
[ ]: accelerator = Accelerator()
model, optimizer, train_loader, val_loader = accelerator.prepare(model,
    ↳optimizer, train_loader, val_loader)

scheduler = ReduceLROnPlateau(optimizer, mode='max', factor=0.1, patience=2)
num_epochs = 30
best_f1_score = 0
best_precision = 0
best_recall = 0
best_model_state = {}
for epoch in range(num_epochs):
    model.train()
    for batch in train_loader:
        input_ids = batch['input_ids']
        labels = batch['labels']

        optimizer.zero_grad()
        logits = model(input_ids)
        loss = criterion(logits.view(-1, NUM_LABELS), labels.view(-1))
        accelerator.backward(loss)
        optimizer.step()

    # Validation
    model.eval()
    true_ner_tags = []
    predicted_ner_tags = []
    with torch.no_grad():
        for batch in val_loader:
            input_ids = batch['input_ids']
            labels = batch['labels']
            logits = model(input_ids)
            predicted_labels = logits.argmax(dim=2)
            true_ner_tags.extend([list(map(tags.__getitem__, label_indices[:
↳length]))) for label_indices, length in zip(labels, batch['input_lengths'])])
            predicted_ner_tags.extend([list(map(tags.__getitem__,
↳label_indices[:length]))) for label_indices, length in zip(predicted_labels,
↳batch['input_lengths'])])

    precision, recall, f1 = evaluate(
```

```

        itertools.chain(*true_ner_tags),
        itertools.chain(*predicted_ner_tags)
    )
    scheduler.step(f1)
    if f1 > best_f1_score:
        best_f1_score = f1
        best_precision = precision
        best_recall = recall
        best_model_state = copy.deepcopy(model.state_dict())
    print(f"Epoch {epoch+1}/{num_epochs}, Loss: {loss.item()}, precision:␣
↪{precision}, recall: {recall}, F1: {f1}")

print(f"Best precision, recall, F1 score is {best_precision}, {best_recall},␣
↪{best_f1_score}")
torch.save({
    'word2idx': word2idx,
    'model_state_dict': best_model_state
}, 'task1.pt')

```

processed 51362 tokens with 5942 phrases; found: 4385 phrases; correct: 2951.
accuracy: 53.77%; (non-0)

accuracy: 91.54%; precision: 67.30%; recall: 49.66%; FB1: 57.15
 LOC: precision: 84.77%; recall: 58.46%; FB1: 69.20 1267
 MISC: precision: 67.09%; recall: 34.71%; FB1: 45.75 477
 ORG: precision: 48.00%; recall: 41.98%; FB1: 44.79 1173
 PER: precision: 67.71%; recall: 53.96%; FB1: 60.06 1468

Epoch 1/30, Loss: 0.3174291253089905, precision: 67.2976054732041, recall:
49.6634129922585, F1: 57.15115716084051

processed 51362 tokens with 5942 phrases; found: 5215 phrases; correct: 3954.
accuracy: 70.46%; (non-0)

accuracy: 94.46%; precision: 75.82%; recall: 66.54%; FB1: 70.88
 LOC: precision: 89.92%; recall: 70.44%; FB1: 79.00 1439
 MISC: precision: 81.38%; recall: 62.58%; FB1: 70.75 709
 ORG: precision: 60.82%; recall: 61.60%; FB1: 61.21 1358
 PER: precision: 73.55%; recall: 68.24%; FB1: 70.80 1709

Epoch 2/30, Loss: 0.14918768405914307, precision: 75.81975071907958, recall:
66.54325143049479, F1: 70.87926862059695

processed 51362 tokens with 5942 phrases; found: 5444 phrases; correct: 4330.
accuracy: 75.69%; (non-0)

accuracy: 95.26%; precision: 79.54%; recall: 72.87%; FB1: 76.06
 LOC: precision: 85.65%; recall: 81.55%; FB1: 83.55 1749
 MISC: precision: 87.19%; recall: 67.90%; FB1: 76.34 718
 ORG: precision: 70.44%; recall: 65.03%; FB1: 67.62 1238
 PER: precision: 76.71%; recall: 72.42%; FB1: 74.50 1739

Epoch 3/30, Loss: 0.12216146290302277, precision: 79.53710506980161, recall:
72.871087176035, F1: 76.05831723168804

processed 51362 tokens with 5942 phrases; found: 5516 phrases; correct: 4477.

accuracy: 78.08%; (non-0)
 accuracy: 95.61%; precision: 81.16%; recall: 75.35%; FB1: 78.15
 LOC: precision: 87.75%; recall: 82.25%; FB1: 84.91 1722
 MISC: precision: 81.74%; recall: 72.34%; FB1: 76.75 816
 ORG: precision: 73.14%; recall: 66.00%; FB1: 69.38 1210
 PER: precision: 79.98%; recall: 76.76%; FB1: 78.34 1768
 Epoch 4/30, Loss: 0.0946226716041565, precision: 81.16388687454678, recall: 75.34500168293505, F1: 78.14627334613371
 processed 51362 tokens with 5942 phrases; found: 5308 phrases; correct: 4391.
 accuracy: 76.22%; (non-0)
 accuracy: 95.59%; precision: 82.72%; recall: 73.90%; FB1: 78.06
 LOC: precision: 90.28%; recall: 81.93%; FB1: 85.90 1667
 MISC: precision: 81.12%; recall: 74.08%; FB1: 77.44 842
 ORG: precision: 74.37%; recall: 65.77%; FB1: 69.81 1186
 PER: precision: 81.90%; recall: 71.72%; FB1: 76.47 1613
 Epoch 5/30, Loss: 0.06255591660737991, precision: 82.72418990203467, recall: 73.89767754964657, F1: 78.06222222222222
 processed 51362 tokens with 5942 phrases; found: 5679 phrases; correct: 4476.
 accuracy: 78.26%; (non-0)
 accuracy: 95.47%; precision: 78.82%; recall: 75.33%; FB1: 77.03
 LOC: precision: 91.22%; recall: 81.49%; FB1: 86.08 1641
 MISC: precision: 85.84%; recall: 71.69%; FB1: 78.13 770
 ORG: precision: 64.96%; recall: 68.31%; FB1: 66.59 1410
 PER: precision: 75.46%; recall: 76.11%; FB1: 75.78 1858
 Epoch 6/30, Loss: 0.019350890070199966, precision: 78.81669307976756, recall: 75.32817233254796, F1: 77.03295757680063
 processed 51362 tokens with 5942 phrases; found: 5486 phrases; correct: 4455.
 accuracy: 77.33%; (non-0)
 accuracy: 95.67%; precision: 81.21%; recall: 74.97%; FB1: 77.97
 LOC: precision: 90.42%; recall: 83.23%; FB1: 86.68 1691
 MISC: precision: 82.69%; recall: 72.56%; FB1: 77.30 809
 ORG: precision: 72.63%; recall: 67.86%; FB1: 70.16 1253
 PER: precision: 77.73%; recall: 73.13%; FB1: 75.36 1733
 Epoch 7/30, Loss: 0.06133827567100525, precision: 81.20670798395918, recall: 74.97475597441938, F1: 77.966398319916
 processed 51362 tokens with 5942 phrases; found: 5639 phrases; correct: 4564.
 accuracy: 79.31%; (non-0)
 accuracy: 95.78%; precision: 80.94%; recall: 76.81%; FB1: 78.82
 LOC: precision: 89.06%; recall: 85.08%; FB1: 87.03 1755
 MISC: precision: 83.40%; recall: 72.99%; FB1: 77.85 807
 ORG: precision: 72.41%; recall: 68.68%; FB1: 70.49 1272
 PER: precision: 77.95%; recall: 76.38%; FB1: 77.16 1805
 Epoch 8/30, Loss: 0.03472890332341194, precision: 80.93633622982799, recall: 76.80915516661057, F1: 78.81875485709354
 processed 51362 tokens with 5942 phrases; found: 5511 phrases; correct: 4539.
 accuracy: 78.75%; (non-0)
 accuracy: 95.87%; precision: 82.36%; recall: 76.39%; FB1: 79.26
 LOC: precision: 90.67%; recall: 84.59%; FB1: 87.52 1714

MISC: precision: 83.75%; recall: 73.21%; FB1: 78.12 806
 ORG: precision: 73.76%; recall: 68.75%; FB1: 71.17 1250
 PER: precision: 79.72%; recall: 75.35%; FB1: 77.48 1741
 Epoch 9/30, Loss: 0.014202786609530449, precision: 82.36254763200871, recall: 76.38842140693369, F1: 79.26307517680956
 processed 51362 tokens with 5942 phrases; found: 5728 phrases; correct: 4579.
 accuracy: 79.75%; (non-0)
 accuracy: 95.79%; precision: 79.94%; recall: 77.06%; FB1: 78.47
 LOC: precision: 88.28%; recall: 85.68%; FB1: 86.96 1783
 MISC: precision: 79.63%; recall: 73.75%; FB1: 76.58 854
 ORG: precision: 70.98%; recall: 69.13%; FB1: 70.04 1306
 PER: precision: 78.32%; recall: 75.90%; FB1: 77.09 1785
 Epoch 10/30, Loss: 0.040178555995225906, precision: 79.94064245810057, recall: 77.0615954224167, F1: 78.47472150814053
 processed 51362 tokens with 5942 phrases; found: 5678 phrases; correct: 4564.
 accuracy: 79.40%; (non-0)
 accuracy: 95.75%; precision: 80.38%; recall: 76.81%; FB1: 78.55
 LOC: precision: 89.08%; recall: 85.30%; FB1: 87.15 1759
 MISC: precision: 80.45%; recall: 73.21%; FB1: 76.66 839
 ORG: precision: 72.35%; recall: 68.68%; FB1: 70.47 1273
 PER: precision: 77.53%; recall: 76.06%; FB1: 76.79 1807
 Epoch 11/30, Loss: 0.020247124135494232, precision: 80.38041563930962, recall: 76.80915516661057, F1: 78.55421686746989
 processed 51362 tokens with 5942 phrases; found: 5713 phrases; correct: 4586.
 accuracy: 79.46%; (non-0)
 accuracy: 95.75%; precision: 80.27%; recall: 77.18%; FB1: 78.70
 LOC: precision: 88.21%; recall: 85.90%; FB1: 87.04 1789
 MISC: precision: 80.96%; recall: 73.32%; FB1: 76.95 835
 ORG: precision: 72.09%; recall: 68.75%; FB1: 70.38 1279
 PER: precision: 77.90%; recall: 76.55%; FB1: 77.22 1810
 Epoch 12/30, Loss: 0.0029119173996150494, precision: 80.27306143882373, recall: 77.17940087512622, F1: 78.6958386958387
 processed 51362 tokens with 5942 phrases; found: 5671 phrases; correct: 4581.
 accuracy: 79.46%; (non-0)
 accuracy: 95.78%; precision: 80.78%; recall: 77.10%; FB1: 78.89
 LOC: precision: 88.67%; recall: 85.63%; FB1: 87.12 1774
 MISC: precision: 82.02%; recall: 73.21%; FB1: 77.36 823
 ORG: precision: 72.78%; recall: 68.98%; FB1: 70.83 1271
 PER: precision: 78.09%; recall: 76.44%; FB1: 77.26 1803
 Epoch 13/30, Loss: 0.026096682995557785, precision: 80.7794039851878, recall: 77.09525412319084, F1: 78.89434254714544
 processed 51362 tokens with 5942 phrases; found: 5686 phrases; correct: 4583.
 accuracy: 79.54%; (non-0)
 accuracy: 95.78%; precision: 80.60%; recall: 77.13%; FB1: 78.83
 LOC: precision: 88.58%; recall: 85.68%; FB1: 87.11 1777
 MISC: precision: 81.92%; recall: 73.21%; FB1: 77.32 824
 ORG: precision: 72.44%; recall: 68.98%; FB1: 70.66 1277
 PER: precision: 77.93%; recall: 76.49%; FB1: 77.21 1808

Epoch 14/30, Loss: 0.010102656669914722, precision: 80.60147731269785, recall: 77.128912823965, F1: 78.82696938424493
processed 51362 tokens with 5942 phrases; found: 5662 phrases; correct: 4582.
accuracy: 79.47%; (non-0)
accuracy: 95.81%; precision: 80.93%; recall: 77.11%; FB1: 78.97
LOC: precision: 88.73%; recall: 85.68%; FB1: 87.18 1774
MISC: precision: 81.76%; recall: 73.43%; FB1: 77.37 828
ORG: precision: 73.54%; recall: 68.61%; FB1: 70.99 1251
PER: precision: 78.00%; recall: 76.60%; FB1: 77.29 1809
Epoch 15/30, Loss: 0.003955557942390442, precision: 80.92546803249735, recall: 77.11208347357793, F1: 78.97276801103067
processed 51362 tokens with 5942 phrases; found: 5673 phrases; correct: 4582.
accuracy: 79.48%; (non-0)
accuracy: 95.80%; precision: 80.77%; recall: 77.11%; FB1: 78.90
LOC: precision: 88.48%; recall: 85.68%; FB1: 87.06 1779
MISC: precision: 81.52%; recall: 73.21%; FB1: 77.14 828
ORG: precision: 73.33%; recall: 68.68%; FB1: 70.93 1256
PER: precision: 78.01%; recall: 76.66%; FB1: 77.33 1810
Epoch 16/30, Loss: 0.01319663506001234, precision: 80.7685527939362, recall: 77.11208347357793, F1: 78.89797675419715
processed 51362 tokens with 5942 phrases; found: 5677 phrases; correct: 4582.
accuracy: 79.48%; (non-0)
accuracy: 95.79%; precision: 80.71%; recall: 77.11%; FB1: 78.87
LOC: precision: 88.53%; recall: 85.68%; FB1: 87.08 1778
MISC: precision: 81.72%; recall: 73.21%; FB1: 77.23 826
ORG: precision: 72.83%; recall: 68.75%; FB1: 70.73 1266
PER: precision: 78.09%; recall: 76.60%; FB1: 77.34 1807
Epoch 17/30, Loss: 0.03489071875810623, precision: 80.71164347366567, recall: 77.11208347357793, F1: 78.87081504432395
processed 51362 tokens with 5942 phrases; found: 5681 phrases; correct: 4584.
accuracy: 79.54%; (non-0)
accuracy: 95.80%; precision: 80.69%; recall: 77.15%; FB1: 78.88
LOC: precision: 88.57%; recall: 85.63%; FB1: 87.07 1776
MISC: precision: 81.82%; recall: 73.21%; FB1: 77.28 825
ORG: precision: 72.70%; recall: 68.90%; FB1: 70.75 1271
PER: precision: 78.05%; recall: 76.66%; FB1: 77.35 1809
Epoch 18/30, Loss: 0.004489194601774216, precision: 80.69001936278825, recall: 77.14574217435207, F1: 78.87808655252516
processed 51362 tokens with 5942 phrases; found: 5681 phrases; correct: 4584.
accuracy: 79.54%; (non-0)
accuracy: 95.80%; precision: 80.69%; recall: 77.15%; FB1: 78.88
LOC: precision: 88.57%; recall: 85.63%; FB1: 87.07 1776
MISC: precision: 81.82%; recall: 73.21%; FB1: 77.28 825
ORG: precision: 72.70%; recall: 68.90%; FB1: 70.75 1271
PER: precision: 78.05%; recall: 76.66%; FB1: 77.35 1809
Epoch 19/30, Loss: 0.0014949982287362218, precision: 80.69001936278825, recall: 77.14574217435207, F1: 78.87808655252516
processed 51362 tokens with 5942 phrases; found: 5679 phrases; correct: 4583.

accuracy: 79.52%; (non-0)
 accuracy: 95.79%; precision: 80.70%; recall: 77.13%; FB1: 78.87
 LOC: precision: 88.57%; recall: 85.63%; FB1: 87.07 1776
 MISC: precision: 81.82%; recall: 73.21%; FB1: 77.28 825
 ORG: precision: 72.68%; recall: 68.83%; FB1: 70.70 1270
 PER: precision: 78.10%; recall: 76.66%; FB1: 77.37 1808
 Epoch 20/30, Loss: 0.010435584001243114, precision: 80.70082761049481, recall: 77.128912823965, F1: 78.87445142414595
 processed 51362 tokens with 5942 phrases; found: 5678 phrases; correct: 4582.
 accuracy: 79.51%; (non-0)
 accuracy: 95.79%; precision: 80.70%; recall: 77.11%; FB1: 78.86
 LOC: precision: 88.57%; recall: 85.63%; FB1: 87.07 1776
 MISC: precision: 81.82%; recall: 73.21%; FB1: 77.28 825
 ORG: precision: 72.66%; recall: 68.75%; FB1: 70.65 1269
 PER: precision: 78.10%; recall: 76.66%; FB1: 77.37 1808
 Epoch 21/30, Loss: 0.011386066675186157, precision: 80.69742867206763, recall: 77.11208347357793, F1: 78.86402753872633
 processed 51362 tokens with 5942 phrases; found: 5678 phrases; correct: 4582.
 accuracy: 79.51%; (non-0)
 accuracy: 95.79%; precision: 80.70%; recall: 77.11%; FB1: 78.86
 LOC: precision: 88.57%; recall: 85.63%; FB1: 87.07 1776
 MISC: precision: 81.82%; recall: 73.21%; FB1: 77.28 825
 ORG: precision: 72.66%; recall: 68.75%; FB1: 70.65 1269
 PER: precision: 78.10%; recall: 76.66%; FB1: 77.37 1808
 Epoch 22/30, Loss: 0.014350972138345242, precision: 80.69742867206763, recall: 77.11208347357793, F1: 78.86402753872633
 processed 51362 tokens with 5942 phrases; found: 5678 phrases; correct: 4582.
 accuracy: 79.51%; (non-0)
 accuracy: 95.79%; precision: 80.70%; recall: 77.11%; FB1: 78.86
 LOC: precision: 88.57%; recall: 85.63%; FB1: 87.07 1776
 MISC: precision: 81.82%; recall: 73.21%; FB1: 77.28 825
 ORG: precision: 72.66%; recall: 68.75%; FB1: 70.65 1269
 PER: precision: 78.10%; recall: 76.66%; FB1: 77.37 1808
 Epoch 23/30, Loss: 0.005108046345412731, precision: 80.69742867206763, recall: 77.11208347357793, F1: 78.86402753872633
 processed 51362 tokens with 5942 phrases; found: 5678 phrases; correct: 4582.
 accuracy: 79.51%; (non-0)
 accuracy: 95.79%; precision: 80.70%; recall: 77.11%; FB1: 78.86
 LOC: precision: 88.57%; recall: 85.63%; FB1: 87.07 1776
 MISC: precision: 81.82%; recall: 73.21%; FB1: 77.28 825
 ORG: precision: 72.66%; recall: 68.75%; FB1: 70.65 1269
 PER: precision: 78.10%; recall: 76.66%; FB1: 77.37 1808
 Epoch 24/30, Loss: 0.005793038755655289, precision: 80.69742867206763, recall: 77.11208347357793, F1: 78.86402753872633
 processed 51362 tokens with 5942 phrases; found: 5678 phrases; correct: 4582.
 accuracy: 79.51%; (non-0)
 accuracy: 95.79%; precision: 80.70%; recall: 77.11%; FB1: 78.86
 LOC: precision: 88.57%; recall: 85.63%; FB1: 87.07 1776

MISC: precision: 81.82%; recall: 73.21%; FB1: 77.28 825
 ORG: precision: 72.66%; recall: 68.75%; FB1: 70.65 1269
 PER: precision: 78.10%; recall: 76.66%; FB1: 77.37 1808
 Epoch 25/30, Loss: 0.002187727950513363, precision: 80.69742867206763, recall: 77.11208347357793, F1: 78.86402753872633
 processed 51362 tokens with 5942 phrases; found: 5678 phrases; correct: 4582.
 accuracy: 79.51%; (non-0)
 accuracy: 95.79%; precision: 80.70%; recall: 77.11%; FB1: 78.86
 LOC: precision: 88.57%; recall: 85.63%; FB1: 87.07 1776
 MISC: precision: 81.82%; recall: 73.21%; FB1: 77.28 825
 ORG: precision: 72.66%; recall: 68.75%; FB1: 70.65 1269
 PER: precision: 78.10%; recall: 76.66%; FB1: 77.37 1808
 Epoch 26/30, Loss: 0.007478665094822645, precision: 80.69742867206763, recall: 77.11208347357793, F1: 78.86402753872633
 processed 51362 tokens with 5942 phrases; found: 5678 phrases; correct: 4582.
 accuracy: 79.51%; (non-0)
 accuracy: 95.79%; precision: 80.70%; recall: 77.11%; FB1: 78.86
 LOC: precision: 88.57%; recall: 85.63%; FB1: 87.07 1776
 MISC: precision: 81.82%; recall: 73.21%; FB1: 77.28 825
 ORG: precision: 72.66%; recall: 68.75%; FB1: 70.65 1269
 PER: precision: 78.10%; recall: 76.66%; FB1: 77.37 1808
 Epoch 27/30, Loss: 0.010278786532580853, precision: 80.69742867206763, recall: 77.11208347357793, F1: 78.86402753872633
 processed 51362 tokens with 5942 phrases; found: 5678 phrases; correct: 4582.
 accuracy: 79.51%; (non-0)
 accuracy: 95.79%; precision: 80.70%; recall: 77.11%; FB1: 78.86
 LOC: precision: 88.57%; recall: 85.63%; FB1: 87.07 1776
 MISC: precision: 81.82%; recall: 73.21%; FB1: 77.28 825
 ORG: precision: 72.66%; recall: 68.75%; FB1: 70.65 1269
 PER: precision: 78.10%; recall: 76.66%; FB1: 77.37 1808
 Epoch 28/30, Loss: 0.0034595143515616655, precision: 80.69742867206763, recall: 77.11208347357793, F1: 78.86402753872633
 processed 51362 tokens with 5942 phrases; found: 5678 phrases; correct: 4582.
 accuracy: 79.51%; (non-0)
 accuracy: 95.79%; precision: 80.70%; recall: 77.11%; FB1: 78.86
 LOC: precision: 88.57%; recall: 85.63%; FB1: 87.07 1776
 MISC: precision: 81.82%; recall: 73.21%; FB1: 77.28 825
 ORG: precision: 72.66%; recall: 68.75%; FB1: 70.65 1269
 PER: precision: 78.10%; recall: 76.66%; FB1: 77.37 1808
 Epoch 29/30, Loss: 0.00442464929074049, precision: 80.69742867206763, recall: 77.11208347357793, F1: 78.86402753872633
 processed 51362 tokens with 5942 phrases; found: 5678 phrases; correct: 4582.
 accuracy: 79.51%; (non-0)
 accuracy: 95.79%; precision: 80.70%; recall: 77.11%; FB1: 78.86
 LOC: precision: 88.57%; recall: 85.63%; FB1: 87.07 1776
 MISC: precision: 81.82%; recall: 73.21%; FB1: 77.28 825
 ORG: precision: 72.66%; recall: 68.75%; FB1: 70.65 1269
 PER: precision: 78.10%; recall: 76.66%; FB1: 77.37 1808

Epoch 30/30, Loss: 0.0007846411899663508, precision: 80.69742867206763, recall: 77.11208347357793, F1: 78.86402753872633
Best precision, recall, F1 score is 82.36254763200871, 76.38842140693369, 79.26307517680956

```
[ ]: !wget http://nlp.stanford.edu/data/glove.6B.zip
      !unzip glove.6B.zip
```

```
--2023-11-11 00:56:49-- http://nlp.stanford.edu/data/glove.6B.zip
Resolving nlp.stanford.edu (nlp.stanford.edu)... 171.64.67.140
Connecting to nlp.stanford.edu (nlp.stanford.edu)|171.64.67.140|:80...
connected.
HTTP request sent, awaiting response... 302 Found
Location: https://nlp.stanford.edu/data/glove.6B.zip [following]
--2023-11-11 00:56:49-- https://nlp.stanford.edu/data/glove.6B.zip
Connecting to nlp.stanford.edu (nlp.stanford.edu)|171.64.67.140|:443...
connected.
HTTP request sent, awaiting response... 301 Moved Permanently
Location: https://downloads.cs.stanford.edu/nlp/data/glove.6B.zip [following]
--2023-11-11 00:56:50-- https://downloads.cs.stanford.edu/nlp/data/glove.6B.zip
Resolving downloads.cs.stanford.edu (downloads.cs.stanford.edu)... 171.64.64.22
Connecting to downloads.cs.stanford.edu
(downloads.cs.stanford.edu)|171.64.64.22|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 862182613 (822M) [application/zip]
Saving to: 'glove.6B.zip'
```

```
glove.6B.zip      100%[=====>] 822.24M  5.01MB/s   in 2m 39s
```

```
2023-11-11 00:59:30 (5.16 MB/s) - 'glove.6B.zip' saved [862182613/862182613]
```

```
Archive: glove.6B.zip
  inflating: glove.6B.50d.txt
  inflating: glove.6B.100d.txt
  inflating: glove.6B.200d.txt
  inflating: glove.6B.300d.txt
```

```
[ ]: # adjust this to point to location of glove embeddings
      GLOVE_EMBEDDINGS_PATH = './glove.6B.100d.txt'
```

```
[ ]: vocab, embeddings = [], []
      with open('glove.6B.100d.txt', 'rt') as fi:
          full_content = fi.read().strip().split('\n')
      for i in range(len(full_content)):
          i_word = full_content[i].split(' ')[0]
          i_embeddings = [float(val) for val in full_content[i].split(' ')[1:]]
          vocab.append(i_word)
```

```

        embeddings.append(i_embeddings)

vocab_npa = np.array(vocab)
embs_npa = np.array(embeddings)

vocab_npa = np.insert(vocab_npa, 0, PAD)
vocab_npa = np.insert(vocab_npa, 1, UNKNOWN)

pad_emb_npa = np.zeros((1, embs_npa.shape[1]))
unk_emb_npa = np.mean(embs_npa, axis=0, keepdims=True)

embs_npa = np.vstack((pad_emb_npa, unk_emb_npa, embs_npa))

glove_embedding_layer = torch.nn.Embedding.from_pretrained(torch.
    ↪from_numpy(embs_npa).float())

```

2.2 Task 2 training

```

[ ]: word2idx_glove = {
    word: index
    for index, word in enumerate(vocab_npa)
}

```

```

[ ]: def preprocess_sample_glove(sample):
    sample['input_ids'] = [word2idx_glove.get(token.lower(),
    ↪word2idx_glove[UNKNOWN]) for token in sample['tokens']]

    del sample['pos_tags']
    del sample['chunk_tags']
    del sample['tokens']
    del sample['id']

    sample['labels'] = sample['ner_tags']
    del sample['ner_tags']

    return sample
dataset_task2 = dataset.map(preprocess_sample_glove)

```

```
Map:   0%|          | 0/14041 [00:00<?, ? examples/s]
```

```
Map:   0%|          | 0/3250 [00:00<?, ? examples/s]
```

```
Map:   0%|          | 0/3453 [00:00<?, ? examples/s]
```

```

[ ]: train_dataset = dataset_task2['train']
    val_dataset = dataset_task2['validation']

```

```

train_loader = DataLoader(train_dataset, batch_size=batch_size, shuffle=True,
    ↪collate_fn=collate_fn)
val_loader = DataLoader(val_dataset, batch_size=batch_size,
    ↪collate_fn=collate_fn)

```

```

[ ]: model = BiLSTMModel(EMBEDDING_DIMENSION, glove_embedding_layer,
    ↪LSTM_HIDDEN_DIMENSION, NUM_LSTM_LAYER, LSTM_DROPOUT, LINEAR_OUTPUT_DIMENSION,
    ↪NUM_LABELS)
criterion = nn.CrossEntropyLoss(ignore_index=tag2idx[PAD])
optimizer = torch.optim.AdamW(model.parameters(), lr=learning_rate)

```

```

[ ]: accelerator = Accelerator()
model, optimizer, train_loader, val_loader = accelerator.prepare(model,
    ↪optimizer, train_loader, val_loader)

scheduler = ReduceLROnPlateau(optimizer, mode='max', factor=0.1, patience=2)
num_epochs = 20
best_f1_score = 0
best_precision = 0
best_recall = 0
best_model_state = {}
for epoch in range(num_epochs):
    model.train()
    for batch in train_loader:
        input_ids = batch['input_ids']
        labels = batch['labels']

        optimizer.zero_grad()
        logits = model(input_ids)
        loss = criterion(logits.view(-1, NUM_LABELS), labels.view(-1))
        accelerator.backward(loss)
        optimizer.step()

    # Validation
    model.eval()
    true_ner_tags = []
    predicted_ner_tags = []
    with torch.no_grad():
        for batch in val_loader:
            input_ids = batch['input_ids']
            labels = batch['labels']
            logits = model(input_ids)
            predicted_labels = logits.argmax(dim=2)
            true_ner_tags.extend([list(map(tags.__getitem__, label_indices[:
    ↪length]))) for label_indices, length in zip(labels, batch['input_lengths'])])

```



```

        predicted_ner_tags.extend([list(map(tags.__getitem__,
↪label_indices[:length])) for label_indices, length in zip(predicted_labels,
↪batch['input_lengths'])])

    precision, recall, f1 = evaluate(
        itertools.chain(*true_ner_tags),
        itertools.chain(*predicted_ner_tags)
    )
    scheduler.step(f1)
    if f1 > best_f1_score:
        best_f1_score = f1
        best_precision = precision
        best_recall = recall
        best_model_state = copy.deepcopy(model.state_dict())
    print(f"Epoch {epoch+1}/{num_epochs}, Loss: {loss.item()}, precision:
↪{precision}, recall: {recall}, F1: {f1}")

print(f"Best precision, recall, F1 score is {best_precision}, {best_recall},
↪{best_f1_score}")
torch.save({
    'word2idx_glove': word2idx_glove,
    'glove_model_dict': best_model_state
}, 'task2.pt')

```

processed 51362 tokens with 5942 phrases; found: 5731 phrases; correct: 4726.

accuracy: 79.54%; (non-0)

accuracy: 96.13%; precision: 82.46%; recall: 79.54%; FB1: 80.97

LOC: precision: 87.39%; recall: 84.10%; FB1: 85.71 1768

MISC: precision: 72.80%; recall: 64.75%; FB1: 68.54 820

ORG: precision: 71.99%; recall: 68.61%; FB1: 70.26 1278

PER: precision: 89.22%; recall: 90.34%; FB1: 89.78 1865

Epoch 1/20, Loss: 0.20355863869190216, precision: 82.46379340429245, recall: 79.53550992931673, F1: 80.97318598475113

processed 51362 tokens with 5942 phrases; found: 5948 phrases; correct: 5039.

accuracy: 85.21%; (non-0)

accuracy: 96.90%; precision: 84.72%; recall: 84.80%; FB1: 84.76

LOC: precision: 89.64%; recall: 89.98%; FB1: 89.81 1844

MISC: precision: 81.83%; recall: 70.82%; FB1: 75.93 798

ORG: precision: 73.01%; recall: 74.42%; FB1: 73.71 1367

PER: precision: 89.48%; recall: 94.19%; FB1: 91.77 1939

Epoch 2/20, Loss: 0.33997663855552673, precision: 84.71755211835911, recall: 84.80309660047122, F1: 84.76030277544153

processed 51362 tokens with 5942 phrases; found: 5942 phrases; correct: 5161.

accuracy: 86.69%; (non-0)

accuracy: 97.23%; precision: 86.86%; recall: 86.86%; FB1: 86.86

LOC: precision: 89.16%; recall: 93.09%; FB1: 91.08 1918

MISC: precision: 81.76%; recall: 74.40%; FB1: 77.91 839

ORG: precision: 78.23%; recall: 77.18%; FB1: 77.70 1323
 PER: precision: 92.91%; recall: 93.92%; FB1: 93.41 1862
 Epoch 3/20, Loss: 0.11293311417102814, precision: 86.85627734769437, recall: 86.85627734769437, F1: 86.85627734769437
 processed 51362 tokens with 5942 phrases; found: 5794 phrases; correct: 5130.
 accuracy: 85.98%; (non-0)
 accuracy: 97.40%; precision: 88.54%; recall: 86.33%; FB1: 87.42
 LOC: precision: 92.91%; recall: 92.00%; FB1: 92.45 1819
 MISC: precision: 85.79%; recall: 73.97%; FB1: 79.44 795
 ORG: precision: 77.91%; recall: 77.85%; FB1: 77.88 1340
 PER: precision: 93.15%; recall: 93.05%; FB1: 93.10 1840
 Epoch 4/20, Loss: 0.08061516284942627, precision: 88.53986882982396, recall: 86.33456748569506, F1: 87.42331288343557
 processed 51362 tokens with 5942 phrases; found: 5932 phrases; correct: 5230.
 accuracy: 87.75%; (non-0)
 accuracy: 97.53%; precision: 88.17%; recall: 88.02%; FB1: 88.09
 LOC: precision: 89.79%; recall: 94.28%; FB1: 91.98 1929
 MISC: precision: 84.72%; recall: 78.20%; FB1: 81.33 851
 ORG: precision: 80.54%; recall: 78.08%; FB1: 79.29 1300
 PER: precision: 93.41%; recall: 93.92%; FB1: 93.67 1852
 Epoch 5/20, Loss: 0.05059090256690979, precision: 88.16587997302764, recall: 88.01750252440256, F1: 88.09162876873843
 processed 51362 tokens with 5942 phrases; found: 5928 phrases; correct: 5238.
 accuracy: 88.17%; (non-0)
 accuracy: 97.60%; precision: 88.36%; recall: 88.15%; FB1: 88.26
 LOC: precision: 93.56%; recall: 92.60%; FB1: 93.08 1818
 MISC: precision: 80.71%; recall: 78.96%; FB1: 79.82 902
 ORG: precision: 78.66%; recall: 79.42%; FB1: 79.04 1354
 PER: precision: 94.07%; recall: 94.68%; FB1: 94.37 1854
 Epoch 6/20, Loss: 0.06662113219499588, precision: 88.36032388663968, recall: 88.15213732749916, F1: 88.25610783487784
 processed 51362 tokens with 5942 phrases; found: 5997 phrases; correct: 5275.
 accuracy: 89.25%; (non-0)
 accuracy: 97.63%; precision: 87.96%; recall: 88.77%; FB1: 88.37
 LOC: precision: 90.95%; recall: 94.61%; FB1: 92.74 1911
 MISC: precision: 80.68%; recall: 77.44%; FB1: 79.03 885
 ORG: precision: 81.56%; recall: 80.46%; FB1: 81.01 1323
 PER: precision: 92.86%; recall: 94.68%; FB1: 93.76 1878
 Epoch 7/20, Loss: 0.05570973455905914, precision: 87.96064699016175, recall: 88.77482329182094, F1: 88.36585978725186
 processed 51362 tokens with 5942 phrases; found: 6016 phrases; correct: 5297.
 accuracy: 88.95%; (non-0)
 accuracy: 97.67%; precision: 88.05%; recall: 89.15%; FB1: 88.59
 LOC: precision: 92.68%; recall: 93.69%; FB1: 93.18 1857
 MISC: precision: 83.01%; recall: 79.50%; FB1: 81.22 883
 ORG: precision: 78.79%; recall: 80.91%; FB1: 79.84 1377
 PER: precision: 92.58%; recall: 95.44%; FB1: 93.99 1899
 Epoch 8/20, Loss: 0.06039019674062729, precision: 88.04853723404256, recall:

89.14506900033659, F1: 88.5934102692758
processed 51362 tokens with 5942 phrases; found: 5968 phrases; correct: 5271.
accuracy: 88.72%; (non-0)
accuracy: 97.58%; precision: 88.32%; recall: 88.71%; FB1: 88.51
LOC: precision: 91.21%; recall: 94.34%; FB1: 92.75 1900
MISC: precision: 84.42%; recall: 77.55%; FB1: 80.84 847
ORG: precision: 79.54%; recall: 82.33%; FB1: 80.91 1388
PER: precision: 93.78%; recall: 93.32%; FB1: 93.55 1833
Epoch 9/20, Loss: 0.04367757961153984, precision: 88.32104557640751, recall: 88.70750589027263, F1: 88.5138539042821
processed 51362 tokens with 5942 phrases; found: 5857 phrases; correct: 5241.
accuracy: 87.61%; (non-0)
accuracy: 97.63%; precision: 89.48%; recall: 88.20%; FB1: 88.84
LOC: precision: 93.90%; recall: 92.98%; FB1: 93.44 1819
MISC: precision: 83.31%; recall: 79.61%; FB1: 81.42 881
ORG: precision: 83.24%; recall: 78.15%; FB1: 80.62 1259
PER: precision: 92.26%; recall: 95.06%; FB1: 93.64 1898
Epoch 10/20, Loss: 0.018665170297026634, precision: 89.48267030903193, recall: 88.20262537866039, F1: 88.83803712178998
processed 51362 tokens with 5942 phrases; found: 5968 phrases; correct: 5290.
accuracy: 89.31%; (non-0)
accuracy: 97.71%; precision: 88.64%; recall: 89.03%; FB1: 88.83
LOC: precision: 93.28%; recall: 92.92%; FB1: 93.10 1830
MISC: precision: 85.28%; recall: 79.83%; FB1: 82.46 863
ORG: precision: 79.58%; recall: 82.55%; FB1: 81.04 1391
PER: precision: 92.36%; recall: 94.46%; FB1: 93.40 1884
Epoch 11/20, Loss: 0.020468614995479584, precision: 88.63941018766755, recall: 89.02726354762707, F1: 88.83291351805205
processed 51362 tokens with 5942 phrases; found: 5892 phrases; correct: 5280.
accuracy: 88.47%; (non-0)
accuracy: 97.67%; precision: 89.61%; recall: 88.86%; FB1: 89.23
LOC: precision: 93.32%; recall: 93.58%; FB1: 93.45 1842
MISC: precision: 84.72%; recall: 79.39%; FB1: 81.97 864
ORG: precision: 82.25%; recall: 81.21%; FB1: 81.73 1324
PER: precision: 93.45%; recall: 94.46%; FB1: 93.95 1862
Epoch 12/20, Loss: 0.03831753507256508, precision: 89.61303462321793, recall: 88.85897004375632, F1: 89.23440932905189
processed 51362 tokens with 5942 phrases; found: 5952 phrases; correct: 5279.
accuracy: 88.45%; (non-0)
accuracy: 97.59%; precision: 88.69%; recall: 88.84%; FB1: 88.77
LOC: precision: 92.65%; recall: 94.07%; FB1: 93.35 1865
MISC: precision: 83.43%; recall: 79.72%; FB1: 81.53 881
ORG: precision: 81.01%; recall: 80.46%; FB1: 80.73 1332
PER: precision: 92.69%; recall: 94.30%; FB1: 93.49 1874
Epoch 13/20, Loss: 0.010882445611059666, precision: 88.69287634408603, recall: 88.84214069336923, F1: 88.76744577097696
processed 51362 tokens with 5942 phrases; found: 5939 phrases; correct: 5268.
accuracy: 88.54%; (non-0)

accuracy: 97.60%; precision: 88.70%; recall: 88.66%; FB1: 88.68
 LOC: precision: 91.32%; recall: 93.96%; FB1: 92.62 1890
 MISC: precision: 85.70%; recall: 79.28%; FB1: 82.37 853
 ORG: precision: 81.21%; recall: 78.97%; FB1: 80.08 1304
 PER: precision: 92.60%; recall: 95.11%; FB1: 93.84 1892
 Epoch 14/20, Loss: 0.023185444995760918, precision: 88.70180165010945, recall: 88.6570178391114, F1: 88.67940409056476
 processed 51362 tokens with 5942 phrases; found: 5942 phrases; correct: 5294.
 accuracy: 88.20%; (non-0)
 accuracy: 97.67%; precision: 89.09%; recall: 89.09%; FB1: 89.09
 LOC: precision: 94.21%; recall: 92.98%; FB1: 93.59 1813
 MISC: precision: 83.84%; recall: 80.48%; FB1: 82.13 885
 ORG: precision: 81.96%; recall: 81.66%; FB1: 81.81 1336
 PER: precision: 91.67%; recall: 94.95%; FB1: 93.28 1908
 Epoch 15/20, Loss: 0.001434061210602522, precision: 89.09458094917537, recall: 89.09458094917537, F1: 89.09458094917537
 processed 51362 tokens with 5942 phrases; found: 5937 phrases; correct: 5303.
 accuracy: 88.89%; (non-0)
 accuracy: 97.75%; precision: 89.32%; recall: 89.25%; FB1: 89.28
 LOC: precision: 92.55%; recall: 93.96%; FB1: 93.25 1865
 MISC: precision: 84.91%; recall: 79.93%; FB1: 82.35 868
 ORG: precision: 83.13%; recall: 80.84%; FB1: 81.97 1304
 PER: precision: 92.42%; recall: 95.33%; FB1: 93.85 1900
 Epoch 16/20, Loss: 0.00433164881542325, precision: 89.32120599629442, recall: 89.24604510265904, F1: 89.28360973145888
 processed 51362 tokens with 5942 phrases; found: 5956 phrases; correct: 5317.
 accuracy: 89.18%; (non-0)
 accuracy: 97.78%; precision: 89.27%; recall: 89.48%; FB1: 89.38
 LOC: precision: 93.01%; recall: 93.47%; FB1: 93.24 1846
 MISC: precision: 84.16%; recall: 80.69%; FB1: 82.39 884
 ORG: precision: 82.29%; recall: 82.10%; FB1: 82.19 1338
 PER: precision: 92.96%; recall: 95.28%; FB1: 94.10 1888
 Epoch 17/20, Loss: 0.002030858537182212, precision: 89.27132303559436, recall: 89.48165600807809, F1: 89.37636577576062
 processed 51362 tokens with 5942 phrases; found: 5962 phrases; correct: 5316.
 accuracy: 89.14%; (non-0)
 accuracy: 97.76%; precision: 89.16%; recall: 89.46%; FB1: 89.31
 LOC: precision: 92.48%; recall: 93.69%; FB1: 93.08 1861
 MISC: precision: 84.72%; recall: 80.59%; FB1: 82.60 877
 ORG: precision: 82.34%; recall: 81.73%; FB1: 82.04 1331
 PER: precision: 92.76%; recall: 95.33%; FB1: 94.03 1893
 Epoch 18/20, Loss: 0.0009912765817716718, precision: 89.16470982891647, recall: 89.464826657691, F1: 89.31451612903226
 processed 51362 tokens with 5942 phrases; found: 5929 phrases; correct: 5299.
 accuracy: 88.85%; (non-0)
 accuracy: 97.75%; precision: 89.37%; recall: 89.18%; FB1: 89.28
 LOC: precision: 92.81%; recall: 93.47%; FB1: 93.14 1850
 MISC: precision: 84.83%; recall: 80.04%; FB1: 82.37 870

ORG: precision: 82.89%; recall: 81.66%; FB1: 82.27 1321
PER: precision: 92.64%; recall: 94.95%; FB1: 93.78 1888
Epoch 19/20, Loss: 0.009812542237341404, precision: 89.37426210153482, recall:
89.17872770111074, F1: 89.27638783590261
processed 51362 tokens with 5942 phrases; found: 5977 phrases; correct: 5316.
accuracy: 89.19%; (non-0)
accuracy: 97.74%; precision: 88.94%; recall: 89.46%; FB1: 89.20
LOC: precision: 92.69%; recall: 93.85%; FB1: 93.26 1860
MISC: precision: 83.47%; recall: 79.93%; FB1: 81.66 883
ORG: precision: 82.19%; recall: 81.88%; FB1: 82.03 1336
PER: precision: 92.57%; recall: 95.39%; FB1: 93.96 1898
Epoch 20/20, Loss: 0.0019614400807768106, precision: 88.94094027103898, recall:
89.464826657691, F1: 89.20211427133148
Best precision, recall, F1 score is 89.27132303559436, 89.48165600807809,
89.37636577576062