Deep Learning CS60010

Assignment-2 Report

Saransh Sharma 20CS30065

en_coarse:

Hyperparameter:

```
SEQ_LEN = 30
EMBEDDING_DIM = 100
HIDDEN_DIM    = 500
NUM_EPOCHS    = 10
BATCH_SIZE    = 10
```

Experiment Setup:

- Dataset Description: The dataset contains a list of sentences wherein the line next to # id marks the beginning of a sentence. Every sentence has an ID which is followed by a list of words and corresponding tags separated by _ _ separator. The tags are annotated using IOB format (I: inside, B: beginning, O: outside). 6 coarse tags exist which are further subdivided into 36 fine tags. The dataset is a fine-grained tagset.

- The number of unique tags in coarse grain setting: 14
- The number of unique words in coarse grain setting: 242153
- Training dataset size: 16778
- Validation dataset size: 871
- Test dataset size: 249980

Performance:

The training loop runs for **4** epochs with a total of 26.6 million parameters and a test loss of **0.29478907585144043.** I and B versions of all 6 tags exist along with no named tag i.e (O tag without any class).

	precision	recall	f1-score	support
<PAD>	1.00	1.00	1.00	3727316
B-CreativeWorks	0.51	0.41	0.45	62124
B-Group	0.48	0.45	0.47	60026
B-Location	0.66	0.56	0.60	67893
B-Medical	0.44	0.24	0.31	22490
B-Person	0.78	0.70	0.74	137666
B-Product	0.33	0.17	0.23	27574
I-CreativeWorks	0.64	0.50	0.57	107463
I-Group	0.58	0.55	0.57	74136
I-Location	0.71	0.66	0.68	63007
I-Medical	0.52	0.29	0.37	10613
I-Person	0.82	0.70	0.76	153751
I-Product	0.40	0.13	0.19	17503
O	0.92	0.97	0.94	2967838
accuracy			0.94	7499400
macro avg	0.63	0.52	0.56	7499400
weighted avg	0.93	0.94	0.94	7499400

Link:
https://drive.google.com/file/d/1PpUzyJWFs5c-CJ11x4nreNXoz6HnHQQR/view?usp=share_link

en_fine:

Hyperparameter:

SEQ_LEN = 30
EMBEDDING_DIM = 100
HIDDEN_DIM = 500
NUM_EPOCHS = 10
BATCH_SIZE = 10

Experiment Setup:

- Dataset Description: The dataset contains a list of sentences wherein the line next to # id marks the beginning of a sentence. Every sentence has an ID which is followed by a list of words and corresponding tags separated by _ _ separator. The tags are annotated using IOB format (I: inside, B: beginning, O: outside). 6 coarse tags exist which are further subdivided into 36 fine tags. The dataset is a fine-grained tagset.

- The number of unique tags in fine grain setting: 68
- The number of unique words in fine grain setting: 242153
- Training dataset size: 16778
- Validation dataset size: 871
- Test dataset size: 249980

Performance:

The model with 26.7 million parameters runs for **5** epochs with a test loss of **0.4470483958721161.** The dataset has 68 tags indicating the absence of several tags from datasets like OtherCORP, TechCORP, etc.

	precision	recall	f1-score	support
<PAD>	1.00	1.00	1.00	3727316
B-AerospaceManufacturer	0.36	0.32	0.34	1015
B-AnatomicalStructure	0.42	0.20	0.27	5838
B-ArtWork	0.17	0.17	0.17	1270
B-Artist	0.51	0.58	0.54	57034
B-Athlete	0.51	0.42	0.46	27624
B-CarManufacturer	0.34	0.25	0.28	2984
B-Cleric	0.36	0.20	0.26	4732
B-Clothing	0.17	0.11	0.14	2243
B-Disease	0.39	0.25	0.30	5622
B-Drink	0.27	0.17	0.20	2246
B-Facility	0.43	0.35	0.39	16181
B-Food	0.13	0.11	0.12	5317
B-HumanSettlement	0.63	0.61	0.62	41099
B-MedicalProcedure	0.32	0.10	0.15	3850
B-Medication/Vaccine	0.22	0.12	0.16	5421
B-MusicalGRP	0.27	0.28	0.27	12969
B-MusicalWork	0.44	0.44	0.44	15303
B-ORG	0.38	0.39	0.39	22414
B-OtherLOC	0.49	0.33	0.39	4635
B-OtherPER	0.28	0.23	0.25	22027
B-OtherPROD	0.19	0.17	0.18	11833
B-Politician	0.25	0.25	0.25	15990
B-PrivateCorp	0.18	0.29	0.22	810
B-PublicCorp	0.27	0.22	0.24	6825
B-Scientist	0.14	0.07	0.09	4928
B-Software	0.36	0.24	0.29	8962
B-SportsGRP	0.60	0.50	0.55	13009
B-SportsManager	0.34	0.25	0.29	5331
B-Station	0.64	0.47	0.54	5978
B-Symptom	0.38	0.24	0.29	1759
B-Vehicle	0.30	0.12	0.17	5935
B-VisualWork	0.43	0.30	0.35	19677
B-WrittenWork	0.52	0.29	0.37	16912
I-AerospaceManufacturer	0.44	0.44	0.44	802
I-AnatomicalStructure	0.45	0.23	0.31	2152
I-ArtWork	0.32	0.17	0.22	2966
I-Artist	0.54	0.55	0.54	59256
I-Athlete	0.54	0.39	0.45	28000
I-CarManufacturer	0.56	0.27	0.36	1192
I-Cleric	0.47	0.25	0.32	6381
I-Clothing	0.28	0.17	0.21	831
I-Disease	0.41	0.33	0.36	3892
I-Drink	0.21	0.08	0.11	806
I-Facility	0.53	0.47	0.49	24859
I-Food	0.19	0.07	0.11	1930
I-HumanSettlement	0.78	0.67	0.72	19317
I-MedicalProcedure	0.32	0.22	0.26	2590
I-Medication/Vaccine	0.30	0.13	0.18	1166
I-MusicalGRP	0.37	0.30	0.33	14675
I-MusicalWork	0.61	0.46	0.52	30093
I-ORG	0.54	0.47	0.50	34336
I-OtherLOC	0.61	0.47	0.53	9867
I-OtherPER	0.32	0.27	0.29	27825
I-OtherPROD	0.25	0.16	0.19	8983
I-Politician	0.31	0.30	0.30	20957
I-PrivateCorp	0.25	0.34	0.29	757
I-PublicCorp	0.36	0.17	0.23	4689
I-Scientist	0.14	0.11	0.13	5838
I-Software	0.42	0.27	0.33	9648
I-SportsGRP	0.67	0.59	0.62	17685
I-SportsManager	0.39	0.24	0.30	5494
I-Station	0.70	0.66	0.68	8964
I-Symptom	0.39	0.28	0.33	813
I-Vehicle	0.26	0.13	0.17	4953
I-VisualWork	0.52	0.29	0.37	35654
I-WrittenWork	0.70	0.44	0.54	29102
O	0.92	0.97	0.95	2967838
accuracy			0.92	7499400
macro avg	0.41	0.31	0.35	7499400
weighted avg	0.91	0.92	0.92	7499400

Link:
https://drive.google.com/file/d/1-07zrEDTzKlIFbv67VrF7vRlVeNO8wVC/view?usp=share_link

bn_coarse:

Hyperparameter:

SEQ_LEN = 30

EMBEDDING_DIM = 100
HIDDEN_DIM = 500
NUM_EPOCHS = 10
BATCH_SIZE = 10

Experiment Setup:

- Dataset Description: The dataset contains a list of sentences wherein the line next to # id marks the beginning of a sentence. Every sentence has an ID which is followed by a list of words and corresponding tags separated by _ _ separator. The tags are annotated using IOB format (I: inside, B: beginning, O: outside). 6 coarse tags exist which are further subdivided into 36 fine tags. The dataset is a fine-grained tagset.

- The number of unique words in coarse grain setting: 42617
- The number of unique tags in coarse grain setting: 14
- Training dataset size: 9708
- Validation dataset size: 507
- Test dataset size: 19859

Performance:

The training loop runs for **5** epochs with a total of 6.7 million parameters and a test loss of **0.1843109428882599.** I and B versions of all 6 tags exist along with no named tag i.e (O tag without any class).

	precision	recall	f1-score	support
<PAD>	1.00	1.00	1.00	339381
B-CreativeWorks	0.57	0.53	0.55	3640
B-Group	0.79	0.69	0.73	3651
B-Location	0.79	0.63	0.70	7375
B-Medical	0.70	0.62	0.66	1919
B-Person	0.71	0.67	0.69	6935
B-Product	0.57	0.44	0.50	1493
I-CreativeWorks	0.75	0.56	0.64	4698
I-Group	0.87	0.79	0.83	4970
I-Location	0.81	0.72	0.76	3302
I-Medical	0.82	0.65	0.72	669
I-Person	0.75	0.70	0.73	7696
I-Product	0.72	0.48	0.58	762
O	0.95	0.98	0.97	209279
accuracy			0.97	595770
macro avg	0.77	0.68	0.72	595770
weighted avg	0.96	0.97	0.96	595770

Link:
https://drive.google.com/file/d/1MRsP3wFF8EMCI8-Z_VHzWuMJs6x1Fdn0/view?usp=share_link

bn_fine:

Hyperparameter:

```
SEQ_LEN = 30
EMBEDDING_DIM = 100
HIDDEN_DIM    = 500
NUM_EPOCHS    = 10
BATCH_SIZE    = 10
```

Experiment Setup:

- Dataset Description: The dataset contains a list of sentences wherein the line next to # id marks the beginning of a sentence. Every sentence has an ID which is followed by a list of words and corresponding tags separated by _ _ separator. The tags are annotated using IOB format (I: inside, B: beginning, O: outside). 6 coarse tags exist which are further subdivided into 36 fine tags. The dataset is a fine-grained tagset.

- The number of unique tags in fine grain setting: 68
- The number of unique words in fine grain setting: 42617
- Training dataset size: 9708
- Validation dataset size: 507
- Test dataset size: 19859

Performance:

The model with 6.7 million parameters runs for **5** epochs with a test loss of **0.258697509765625.** The dataset has 68 tags indicating the absence of several tags from datasets like OtherCORP, TechCORP, etc.

	precision	recall	f1-score	support
<PAD>	1.00	1.00	1.00	339381
B-AerospaceManufacturer	0.16	0.05	0.08	97
B-AnatomicalStructure	0.72	0.47	0.57	532
B-ArtWork	0.17	0.01	0.02	455
B-Artist	0.46	0.41	0.43	2744
B-Athlete	0.38	0.29	0.33	1087
B-CarManufacturer	0.57	0.82	0.68	84
B-Cleric	0.39	0.60	0.48	240
B-Clothing	0.25	0.65	0.36	17
B-Disease	0.78	0.58	0.66	554
B-Drink	0.62	0.82	0.71	120
B-Facility	0.56	0.44	0.50	894
B-Food	0.48	0.37	0.41	453
B-HumanSettlement	0.75	0.67	0.70	6011
B-MedicalProcedure	0.76	0.63	0.69	266
B-Medication/Vaccine	0.65	0.51	0.57	462
B-MusicalGRP	0.54	0.55	0.55	300
B-MusicalWork	0.44	0.42	0.43	226
B-ORG	0.86	0.62	0.72	1988
B-OtherLOC	0.76	0.62	0.69	172
B-OtherPER	0.31	0.32	0.31	1117
B-OtherPROD	0.55	0.34	0.42	704
B-Politician	0.42	0.31	0.35	1294
B-PrivateCorp	0.98	0.79	0.87	127
B-PublicCorp	0.62	0.64	0.63	460
B-Scientist	0.38	0.24	0.29	255
B-Software	0.78	0.63	0.70	812
B-SportsGRP	0.76	0.80	0.78	595
B-SportsManager	0.29	0.37	0.33	198
B-Station	0.75	0.79	0.77	298
B-Symptom	0.58	0.90	0.71	105
B-Vehicle	0.74	0.62	0.68	199
B-VisualWork	0.41	0.37	0.39	923
B-WrittenWork	0.67	0.52	0.59	1224
I-AerospaceManufacturer	0.17	0.01	0.02	114
I-AnatomicalStructure	0.66	0.47	0.55	91
I-ArtWork	0.14	0.01	0.01	832
I-Artist	0.51	0.41	0.45	2893
I-Athlete	0.44	0.30	0.36	1165
I-CarManufacturer	0.88	0.91	0.90	57
I-Cleric	0.64	0.74	0.69	268
I-Clothing	0.83	0.71	0.77	7
I-Disease	0.84	0.60	0.70	231
I-Drink	1.00	0.81	0.89	21
I-Facility	0.71	0.66	0.68	905
I-Food	0.69	0.41	0.51	175
I-HumanSettlement	0.81	0.68	0.74	1672
I-MedicalProcedure	0.93	0.69	0.79	155
I-Medication/Vaccine	0.84	0.61	0.71	175
I-MusicalGRP	0.65	0.62	0.64	239
I-MusicalWork	0.51	0.49	0.50	304
I-ORG	0.87	0.76	0.81	2941
I-OtherLOC	0.62	0.67	0.64	187
I-OtherPER	0.34	0.33	0.34	1334
I-OtherPROD	0.73	0.46	0.57	458
I-Politician	0.54	0.38	0.44	1515
I-PrivateCorp	1.00	0.86	0.92	69
I-PublicCorp	0.72	0.75	0.73	282
I-Scientist	0.41	0.28	0.33	319
I-Software	0.78	0.57	0.66	466
I-SportsGRP	0.86	0.87	0.86	1268
I-SportsManager	0.37	0.46	0.41	202
I-Station	0.84	0.90	0.87	538
I-Symptom	1.00	0.41	0.58	17
I-Vehicle	0.74	0.52	0.61	101
I-VisualWork	0.58	0.44	0.50	1546
I-WrittenWork	0.68	0.67	0.67	1550
O	0.95	0.98	0.96	209279
accuracy			0.96	595770
macro avg	0.63	0.55	0.58	595770
weighted avg	0.95	0.96	0.95	595770

Link:
https://drive.google.com/file/d/1cKYF8wm1M7gWFB486HP8yA-ebXfyRM-F/view?usp=share_link

hi_coarse:

Hyperparameter:

SEQ_LEN = 30
EMBEDDING_DIM = 100
HIDDEN_DIM = 500
NUM_EPOCHS = 10
BATCH_SIZE = 10

Experiment Setup:

- Dataset Description: The dataset contains a list of sentences wherein the line next to # id marks the beginning of a sentence. Every sentence has an ID which is followed by a list of words and corresponding tags separated by _ _ separator. The tags are annotated using IOB format (I: inside, B: beginning, O: outside). 6 (IOB for each i.e. total of 18) coarse tags exist which are further subdivided into 36 fine tags (IOB for each i.e. total of 108). The dataset is a fine-grained tagset.

- Number of unique tags in coarse grain setting: 14
- Number of unique words in coarse grain setting: 30589
- Training dataset size: 9632
- Validation dataset size: 514
- Test dataset size: 18399

Performance:

The training loop runs for **5** epochs with a total of 5.5 million parameters and a test loss of **0.20241768658161163.** I and B versions of all 6 tags exist along with no named tag i.e (O tag without any class).

	precision	recall	f1-score	support
<PAD>	1.00	1.00	1.00	258203
B-CreativeWorks	0.66	0.53	0.59	2802
B-Group	0.76	0.73	0.74	3893
B-Location	0.82	0.59	0.68	7171
B-Medical	0.72	0.60	0.66	1973
B-Person	0.69	0.64	0.67	5732
B-Product	0.50	0.57	0.53	1608
I-CreativeWorks	0.77	0.51	0.62	3908
I-Group	0.87	0.81	0.84	5567
I-Location	0.80	0.67	0.73	3251
I-Medical	0.83	0.59	0.69	800
I-Person	0.77	0.65	0.70	6532
I-Product	0.65	0.53	0.58	755
O	0.96	0.98	0.97	249775
accuracy			0.96	551970
macro avg	0.77	0.67	0.71	551970
weighted avg	0.96	0.96	0.96	551970

Link:
https://drive.google.com/file/d/1K3tEwrwUfqsX7Nbl7snZ8MgU6W8SUuXs/view?usp=share_link

hi_fine:

Hyperparameter:

```
SEQ_LEN = 30
EMBEDDING_DIM = 100
HIDDEN_DIM    = 500
NUM_EPOCHS    = 10
BATCH_SIZE    = 10
```

Experiment Setup:

- Dataset Description: The dataset contains a list of sentences wherein the line next to # id marks the beginning of a sentence. Every sentence has an ID which is followed by a list of words and corresponding tags separated by _ _ separator. The tags are annotated using IOB format (I: inside, B: beginning, O: outside). 6 coarse tags exist which are further subdivided into 36 fine tags. The dataset is a fine-grained tagset.

- Number of unique tags in fine grain setting: 68
- Number of unique words in fine grain setting: 30589
- Training dataset size: 9632
- Validation dataset size: 514
- Test dataset size: 18399

Performance:

The model with 5.5 million parameters runs for **6** epochs with a test loss of **0.28121206164360046.** The dataset has 68 tags indicating the absence of several tags from datasets like OtherCORP, TechCORP, etc.

	precision	recall	f1-score	support
<PAD>	1.00	1.00	1.00	258203
B-AerospaceManufacturer	0.36	0.05	0.08	85
B-AnatomicalStructure	0.75	0.57	0.65	489
B-ArtWork	0.67	0.01	0.02	426
B-Artist	0.55	0.35	0.43	1852
B-Athlete	0.67	0.55	0.61	1171
B-CarManufacturer	0.75	0.84	0.79	146
B-Cleric	0.60	0.80	0.69	188
B-Clothing	0.59	0.75	0.66	77
B-Disease	0.77	0.57	0.66	633
B-Drink	0.73	0.70	0.72	135
B-Facility	0.54	0.41	0.47	859
B-Food	0.46	0.54	0.50	428
B-HumanSettlement	0.73	0.64	0.68	5825
B-MedicalProcedure	0.69	0.63	0.66	334
B-Medication/Vaccine	0.50	0.64	0.56	377
B-MusicalGRP	0.58	0.72	0.64	173
B-MusicalWork	0.28	0.75	0.41	44
B-ORG	0.69	0.68	0.68	1847
B-OtherLOC	0.61	0.64	0.63	231
B-OtherPER	0.31	0.33	0.32	741
B-OtherPROD	0.38	0.39	0.39	778
B-Politician	0.36	0.48	0.41	1155
B-PrivateCorp	0.90	0.68	0.78	84
B-PublicCorp	0.66	0.61	0.63	416
B-Scientist	0.30	0.57	0.39	132
B-Software	0.76	0.69	0.72	701
B-SportsGRP	0.90	0.79	0.84	1142
B-SportsManager	0.31	0.02	0.03	493
B-Station	0.65	0.74	0.69	256
B-Symptom	0.70	0.76	0.73	140
B-Vehicle	0.63	0.69	0.66	190
B-VisualWork	0.56	0.37	0.45	753
B-WrittenWork	0.63	0.60	0.61	878
I-AerospaceManufacturer	1.00	0.03	0.06	94
I-AnatomicalStructure	0.92	0.53	0.67	109
I-ArtWork	0.80	0.00	0.01	1020
I-Artist	0.60	0.40	0.48	1931
I-Athlete	0.72	0.53	0.61	1260
I-CarManufacturer	0.90	0.90	0.90	69
I-Cleric	0.74	0.86	0.79	200
I-Clothing	0.67	0.29	0.40	7
I-Disease	0.83	0.62	0.71	303
I-Drink	0.82	0.49	0.61	63
I-Facility	0.63	0.40	0.49	747
I-Food	0.71	0.62	0.66	151
I-HumanSettlement	0.80	0.67	0.73	1712
I-MedicalProcedure	0.85	0.69	0.76	222
I-Medication/Vaccine	0.82	0.63	0.71	137
I-MusicalGRP	0.80	0.80	0.80	143
I-MusicalWork	0.31	0.76	0.44	29
I-ORG	0.89	0.75	0.81	3046
I-OtherLOC	0.82	0.71	0.76	259
I-OtherPER	0.33	0.41	0.37	887
I-OtherPROD	0.61	0.41	0.49	434
I-Politician	0.49	0.50	0.49	1420
I-PrivateCorp	0.95	0.77	0.85	81
I-PublicCorp	0.66	0.68	0.67	252
I-Scientist	0.34	0.57	0.43	175
I-Software	0.81	0.55	0.65	360
I-SportsGRP	0.92	0.78	0.85	1882
I-SportsManager	0.28	0.01	0.02	659
I-Station	0.90	0.77	0.83	533
I-Symptom	0.81	0.59	0.68	29
I-Vehicle	0.76	0.78	0.77	100
I-VisualWork	0.63	0.49	0.55	1377
I-WrittenWork	0.78	0.73	0.75	1122
O	0.96	0.98	0.97	249775
accuracy			0.96	551970
macro avg	0.67	0.58	0.59	551970
weighted avg	0.95	0.96	0.95	551970

Link:
https://drive.google.com/file/d/1-ErZvIDwuNt6IPHTLzcAkIMvKVm_hdyI/view?usp=share_link