Deep Learning CS60010

Assignment-2 Report

Pranav Mehrotra 20CS10085

Link to all Model Parameters:

https://drive.google.com/drive/folders/178TW4zZBZHLlI7kkRvdKafvFtyXgGCb9?usp=sharing

English

en_coarse:

Hyperparameter:

SEQ_LEN = 33
EMBEDDING_DIM = 100
HIDDEN_DIM = 530
NUM_EPOCHS = 13
BATCH_SIZE = 10

Experiment Setup:

- Dataset Description: The sentences in the dataset are listed with the first line next to the # id designating the start of the sentence. Every sentence begins with an ID, then has a list of words and their related tags, each of which is separated by a separator. The tags have IOB format annotations (I: inside, B: beginning, O: outside). There are six coarse tags, which are separated into 36 fine tags. A fine-grained tagset makes up the dataset.

- The number of unique tags in coarse grain setting: 14
- The number of unique words in coarse grain setting: 242180
- Training dataset size: 16778
- Validation dataset size: 871
- Test dataset size: 249980

Performance:

The training loop runs for **6** epochs with a total of 26.9 million parameters and a test loss of **0.33054500818252563.** I and B versions of all six 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	4476021
B-CreativeWorks	0.52	0.39	0.44	62126
B-Group	0.47	0.46	0.46	60026
B-Location	0.63	0.59	0.61	67899
B-Medical	0.39	0.25	0.30	22491
B-Person	0.80	0.67	0.73	137674
B-Product	0.24	0.23	0.23	27579
I-CreativeWorks	0.66	0.47	0.55	107467
I-Group	0.62	0.51	0.56	74142
I-Location	0.71	0.65	0.68	63019
I-Medical	0.49	0.30	0.37	10614
I-Person	0.83	0.69	0.75	153769
I-Product	0.27	0.20	0.23	17506
O	0.92	0.97	0.94	2969007
accuracy			0.94	8249340
macro avg	0.61	0.53	0.56	8249340
weighted avg	0.94	0.94	0.94	8249340

en_fine:

Hyperparameter:

SEQ_LEN = 33
EMBEDDING_DIM = 100
HIDDEN_DIM = 530
NUM_EPOCHS = 13
BATCH_SIZE = 10

Experiment Setup:

- The number of unique tags in fine grain setting: 68
- The number of unique words in fine grain setting: 242180
- Training dataset size: 16778
- Validation dataset size: 871
- Test dataset size: 249980

Performance:

The model with 27.0 million parameters runs for **6** epochs with a test loss of **0.42844489216804504.** 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	4476021
B-AerospaceManufacturer	0.27	0.43	0.33	1015
B-AnatomicalStructure	0.34	0.26	0.29	5838
B-ArtWork	0.17	0.22	0.19	1270
B-Artist	0.53	0.58	0.55	57034
B-Athlete	0.49	0.49	0.49	27629
B-CarManufacturer	0.31	0.28	0.29	2984
B-Cleric	0.32	0.23	0.27	4732
B-Clothing	0.23	0.09	0.13	2244
B-Disease	0.46	0.21	0.29	5623
B-Drink	0.24	0.19	0.21	2246
B-Facility	0.35	0.41	0.38	16184
B-Food	0.14	0.11	0.13	5317
B-HumanSettlement	0.74	0.53	0.62	41102
B-MedicalProcedure	0.16	0.17	0.16	3850
B-Medication/Vaccine	0.25	0.14	0.18	5421
B-MusicalGRP	0.33	0.24	0.28	12969
B-MusicalWork	0.58	0.39	0.47	15304
B-ORG	0.37	0.41	0.39	22414
B-OtherLOC	0.38	0.33	0.35	4635
B-OtherPER	0.26	0.29	0.27	22028
B-OtherPROD	0.20	0.16	0.18	11837
B-Politician	0.32	0.20	0.25	15990
B-PrivateCorp	0.24	0.31	0.27	810
B-PublicCorp	0.35	0.22	0.27	6825
B-Scientist	0.14	0.10	0.12	4928
B-Software	0.44	0.25	0.32	8962
B-SportsGRP	0.61	0.51	0.55	13009
B-SportsManager	0.46	0.18	0.26	5333
B-Station	0.68	0.45	0.54	5978
B-Symptom	0.38	0.27	0.32	1759
B-Vehicle	0.28	0.13	0.18	5935
B-VisualWork	0.39	0.35	0.37	19678
B-WrittenWork	0.47	0.32	0.38	16912
I-AerospaceManufacturer	0.28	0.55	0.37	802
I-AnatomicalStructure	0.49	0.24	0.32	2152
I-ArtWork	0.23	0.22	0.23	2966
I-Artist	0.56	0.56	0.56	59257
I-Athlete	0.51	0.46	0.48	28014
I-CarManufacturer	0.51	0.28	0.36	1192
I-Cleric	0.43	0.26	0.32	6381
I-Clothing	0.37	0.12	0.18	831
I-Disease	0.51	0.30	0.38	3893
I-Drink	0.19	0.11	0.14	806
I-Facility	0.45	0.52	0.49	24865
I-Food	0.21	0.05	0.07	1930
I-HumanSettlement	0.76	0.71	0.74	19323
I-MedicalProcedure	0.33	0.24	0.28	2590
I-Medication/Vaccine	0.39	0.15	0.22	1166
I-MusicalGRP	0.49	0.26	0.34	14675
I-MusicalWork	0.66	0.46	0.54	30095
I-ORG	0.56	0.47	0.51	34338
I-OtherLOC	0.51	0.46	0.49	9867
I-OtherPER	0.28	0.38	0.32	27827
I-OtherPROD	0.22	0.17	0.19	8986
I-Politician	0.39	0.21	0.27	20957
I-PrivateCorp	0.31	0.38	0.34	757
I-PublicCorp	0.37	0.21	0.26	4689
I-Scientist	0.15	0.08	0.11	5838
I-Software	0.37	0.29	0.32	9648
I-SportsGRP	0.75	0.54	0.63	17689
I-SportsManager	0.37	0.23	0.28	5495
I-Station	0.76	0.64	0.70	8964
I-Symptom	0.39	0.40	0.39	813
I-Vehicle	0.23	0.16	0.18	4953
I-VisualWork	0.41	0.39	0.40	35656
I-WrittenWork	0.64	0.49	0.55	29102
O	0.93	0.96	0.95	2969007
accuracy			0.93	8249340
macro avg	0.41	0.33	0.36	8249340
weighted avg	0.92	0.93	0.92	8249340

Bangla

bn_coarse:

Hyperparameter:

SEQ_LEN = 33
EMBEDDING_DIM = 100
HIDDEN_DIM = 530
NUM_EPOCHS = 13
BATCH_SIZE = 10

Experiment Setup:

- The number of unique words in coarse grain setting: 42620
- 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 **6** epochs with a total of 7 million parameters and a test loss of **0.1950886994600296.** 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	398906
B-CreativeWorks	0.67	0.47	0.55	3640
B-Group	0.78	0.68	0.72	3651
B-Location	0.69	0.69	0.69	7375
B-Medical	0.69	0.64	0.66	1919
B-Person	0.78	0.59	0.67	6935
B-Product	0.63	0.44	0.52	1493
I-CreativeWorks	0.78	0.52	0.63	4698
I-Group	0.87	0.79	0.83	4970
I-Location	0.81	0.71	0.76	3302
I-Medical	0.83	0.60	0.70	669
I-Person	0.79	0.63	0.70	7696
I-Product	0.79	0.46	0.58	762
O	0.94	0.98	0.96	209331
accuracy			0.97	655347
macro avg	0.79	0.66	0.71	655347
weighted avg	0.97	0.97	0.97	655347

bn_fine:

Hyperparameter:

SEQ_LEN = 33
EMBEDDING_DIM = 100
HIDDEN_DIM = 530
NUM_EPOCHS = 13
BATCH_SIZE = 10

Experiment Setup:

- 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 7 million parameters runs for **6** epochs with a test loss of **0.24942559003829956.** 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	398906
B-AerospaceManufacturer	0.11	0.02	0.03	97
B-AnatomicalStructure	0.63	0.52	0.57	532
B-ArtWork	0.19	0.01	0.02	455
B-Artist	0.55	0.34	0.42	2744
B-Athlete	0.37	0.35	0.36	1087
B-CarManufacturer	0.62	0.92	0.74	84
B-Cleric	0.57	0.55	0.56	240
B-Clothing	0.20	0.59	0.30	17
B-Disease	0.77	0.63	0.69	554
B-Drink	0.69	0.81	0.74	120
B-Facility	0.66	0.42	0.52	894
B-Food	0.46	0.30	0.37	453
B-HumanSettlement	0.75	0.66	0.70	6011
B-MedicalProcedure	0.61	0.68	0.64	266
B-Medication/Vaccine	0.65	0.50	0.56	462
B-MusicalGRP	0.57	0.58	0.58	300
B-MusicalWork	0.49	0.42	0.45	226
B-ORG	0.71	0.66	0.68	1988
B-OtherLOC	0.74	0.67	0.70	172
B-OtherPER	0.30	0.28	0.29	1117
B-OtherPROD	0.58	0.33	0.42	704
B-Politician	0.43	0.37	0.40	1294
B-PrivateCorp	0.96	0.87	0.91	127
B-PublicCorp	0.65	0.64	0.64	460
B-Scientist	0.29	0.36	0.32	255
B-Software	0.70	0.62	0.66	812
B-SportsGRP	0.83	0.82	0.82	595
B-SportsManager	0.30	0.48	0.37	198
B-Station	0.76	0.80	0.78	298
B-Symptom	0.78	0.81	0.79	105
B-Vehicle	0.78	0.67	0.72	199
B-VisualWork	0.48	0.40	0.44	923
B-WrittenWork	0.73	0.50	0.59	1224
I-AerospaceManufacturer	0.00	0.00	0.00	114
I-AnatomicalStructure	0.59	0.48	0.53	91
I-ArtWork	0.10	0.01	0.01	832
I-Artist	0.56	0.35	0.43	2893
I-Athlete	0.44	0.34	0.39	1165
I-CarManufacturer	0.79	0.93	0.85	57
I-Cleric	0.67	0.71	0.69	268
I-Clothing	0.67	0.57	0.62	7
I-Disease	0.78	0.64	0.70	231
I-Drink	0.89	0.81	0.85	21
I-Facility	0.76	0.65	0.70	905
I-Food	0.72	0.43	0.54	175
I-HumanSettlement	0.82	0.68	0.74	1672
I-MedicalProcedure	0.85	0.71	0.77	155
I-Medication/Vaccine	0.86	0.66	0.75	175
I-MusicalGRP	0.69	0.65	0.67	239
I-MusicalWork	0.60	0.51	0.55	304
I-ORG	0.83	0.74	0.78	2941
I-OtherLOC	0.74	0.70	0.72	187
I-OtherPER	0.41	0.26	0.32	1334
I-OtherPROD	0.75	0.44	0.55	458
I-Politician	0.45	0.41	0.43	1515
I-PrivateCorp	1.00	0.84	0.91	69
I-PublicCorp	0.61	0.73	0.66	282
I-Scientist	0.31	0.34	0.32	319
I-Software	0.73	0.55	0.63	466
I-SportsGRP	0.78	0.91	0.84	1268
I-SportsManager	0.29	0.50	0.37	202
I-Station	0.81	0.91	0.86	538
I-Symptom	0.78	0.41	0.54	17
I-Vehicle	0.67	0.54	0.60	101
I-VisualWork	0.54	0.50	0.52	1546
I-WrittenWork	0.79	0.65	0.71	1550
O	0.95	0.98	0.96	209331
accuracy			0.96	655347
macro avg	0.62	0.56	0.58	655347
weighted avg	0.96	0.96	0.96	655347

Hindi

hi_coarse:

Hyperparameter:

SEQ_LEN = 33
EMBEDDING_DIM = 100
HIDDEN_DIM = 530
NUM_EPOCHS = 13
BATCH_SIZE = 10

Experiment Setup:

- Number of unique tags in coarse grain setting: 14
- Number of unique words in coarse grain setting: 30616
- Training dataset size: 9632
- Validation dataset size: 514
- Test dataset size: 18399

Performance:

The training loop runs for **6** epochs with a total of 5.8 million parameters and a test loss of **0.1808880865573883.** 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	312309
B-CreativeWorks	0.60	0.55	0.57	2803
B-Group	0.77	0.76	0.77	3895
B-Location	0.71	0.68	0.69	7172
B-Medical	0.78	0.58	0.66	1977
B-Person	0.66	0.67	0.66	5735
B-Product	0.60	0.57	0.58	1611
I-CreativeWorks	0.78	0.50	0.61	3909
I-Group	0.84	0.82	0.83	5571
I-Location	0.78	0.68	0.73	3251
I-Medical	0.84	0.57	0.68	804
I-Person	0.75	0.66	0.71	6534
I-Product	0.76	0.51	0.61	757
O	0.96	0.98	0.97	250839
accuracy			0.97	607167
macro avg	0.77	0.68	0.72	607167
weighted avg	0.97	0.97	0.97	607167

hi_fine:

Hyperparameter:

SEQ_LEN = 33
EMBEDDING_DIM = 100
HIDDEN_DIM = 530
NUM_EPOCHS = 13
BATCH_SIZE = 10

Experiment Setup:

- Number of unique tags in fine grain setting: 68
- Number of unique words in fine grain setting: 30616
- Training dataset size: 9632
- Validation dataset size: 514
- Test dataset size: 18399

Performance:

The model with 5.8 million parameters runs for **6** epochs with a test loss of **0.2640095353126526.** 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	312309
B-AerospaceManufacturer	0.50	0.06	0.11	85
B-AnatomicalStructure	0.76	0.56	0.64	490
B-ArtWork	0.25	0.00	0.01	426
B-Artist	0.55	0.39	0.45	1852
B-Athlete	0.69	0.53	0.60	1173
B-CarManufacturer	0.76	0.79	0.78	146
B-Cleric	0.71	0.77	0.74	189
B-Clothing	0.73	0.74	0.74	77
B-Disease	0.79	0.55	0.65	633
B-Drink	0.62	0.74	0.68	136
B-Facility	0.56	0.39	0.46	859
B-Food	0.70	0.45	0.55	428
B-HumanSettlement	0.81	0.62	0.70	5826
B-MedicalProcedure	0.78	0.56	0.65	336
B-MedicationVaccine	0.78	0.52	0.63	378
B-MusicalGRP	0.54	0.78	0.64	174
B-MusicalWork	0.31	0.73	0.44	44
B-ORG	0.84	0.64	0.73	1847
B-OtherLOC	0.71	0.61	0.65	231
B-OtherPER	0.36	0.36	0.36	741
B-OtherPROD	0.65	0.27	0.39	780
B-Politician	0.46	0.46	0.46	1155
B-PrivateCorp	0.77	0.76	0.77	84
B-PublicCorp	0.64	0.66	0.65	417
B-Scientist	0.39	0.59	0.47	132
B-Software	0.85	0.63	0.73	702
B-SportsGRP	0.88	0.80	0.84	1142
B-SportsManager	0.18	0.02	0.03	493
B-Station	0.79	0.78	0.78	256
B-Symptom	0.82	0.74	0.77	140
B-Vehicle	0.71	0.76	0.73	190
B-VisualWork	0.53	0.33	0.41	753
B-WrittenWork	0.67	0.58	0.62	878
I-AerospaceManufacturer	1.00	0.04	0.08	94
I-AnatomicalStructure	0.83	0.53	0.65	109
I-ArtWork	0.41	0.01	0.02	1020
I-Artist	0.59	0.43	0.49	1931
I-Athlete	0.71	0.53	0.61	1262
I-CarManufacturer	0.92	0.84	0.88	69
I-Cleric	0.85	0.82	0.83	200
I-Clothing	0.40	0.29	0.33	7
I-Disease	0.84	0.55	0.67	304
I-Drink	0.93	0.62	0.74	63
I-Facility	0.64	0.43	0.51	747
I-Food	0.89	0.51	0.65	151
I-HumanSettlement	0.85	0.66	0.74	1712
I-MedicalProcedure	0.93	0.63	0.75	224
I-MedicationVaccine	0.93	0.57	0.70	138
I-MusicalGRP	0.60	0.90	0.72	144
I-MusicalWork	0.28	0.62	0.39	29
I-ORG	0.90	0.74	0.81	3048
I-OtherLOC	0.81	0.71	0.76	259
I-OtherPER	0.45	0.41	0.43	887
I-OtherPROD	0.77	0.32	0.45	436
I-Politician	0.59	0.49	0.54	1420
I-PrivateCorp	0.94	0.74	0.83	81
I-PublicCorp	0.69	0.69	0.69	252
I-Scientist	0.53	0.57	0.55	175
I-Software	0.81	0.48	0.60	360
I-SportsGRP	0.89	0.82	0.85	1883
I-SportsManager	0.24	0.04	0.06	659
I-Station	0.88	0.83	0.86	533
I-Symptom	0.86	0.66	0.75	29
I-Vehicle	0.71	0.79	0.75	100
I-VisualWork	0.66	0.48	0.56	1378
I-WrittenWork	0.80	0.68	0.73	1122
O	0.95	0.99	0.97	250839
accuracy			0.96	607167
macro avg	0.69	0.57	0.60	607167
weighted avg	0.96	0.96	0.96	607167