

CS689A: Computational Linguistics for Indian Languages

CS689A Assignment2

Name: Moinuddin

Roll NO.: 231110030

Solution to Problem 1: give name entity tags to 25 sentences

solved this question on the website. and the result is in CS698_manual_Ner_hi.txt file which is in the submitted .zip file

Solution to Problem 2: Fine-tune IndicBERT and IndicNER

I finetuned both models with different hyperparameters and different size of training data. Code and result is in submitted jupyter notebook.

Solution to Problem 3: NER tagging using chatGPT

solved this question and the result is given in CS698_chatgpt_Ner_hi.txt file submitted in .zip file.

Solution to Problem 4: Double Q-Learning

ChatGPT vs. Manual Tagging Precision, Recall, and Macro-F1 Score

Class	Precision	Recall	F1 Score
O	0.992	0.935	0.963
B-PER	0.550	0.611	0.579
I-PER	0.412	0.778	0.538
B-ORG	0.333	0.750	0.462
I-ORG	0.286	1.000	0.444
B-LOC	0.143	0.500	0.222
B-MISC	0.167	1.000	0.286
I-MISC	0.333	1.000	0.500

Macro F1 Score = 0.499

1 Validation data

2 train

3 Test data

Table 1: Performance Metrics for Different Models on Validation data

Model	Class	Precision	Recall	F1 Score
Indic_Ner_3epoch_batch_8 Macro F1 = 0.023	O	0.875	0.0178	0.0348
	B-PER	0.0	0.0	0.0
	I-PER	0.0526	0.0588	0.0556
	B-ORG	0.0	0.0	0.0
	I-ORG	0.0	0.0	0.0
	B-LOC	0.0377	0.8333	0.0721
	I-LOC	0.0	0.0	0.0
IndicBERT-3epoch_batch_16 Macro F1 = 0.13	O	0.833	0.998	0.908
	B-PER	1.0	0.0	0.0
	I-PER	1.0	0.0	0.0
	B-ORG	0.0	0.0	0.0
	I-ORG	1.0	0.0	0.0
	B-LOC	0.0	0.0	0.0
	I-LOC	1.0	0.0	0.0
Indic_Ner_3epoch_batch_16 Macro F1 = 0.023	O	0.875	0.0178	0.0348
	B-PER	0.0	0.0	0.0
	I-PER	0.0526	0.0588	0.0556
	B-ORG	0.0	0.0	0.0
	I-ORG	0.0	0.0	0.0
	B-LOC	0.0405	0.8889	0.0775
	I-LOC	0.0	0.0	0.0
Indic_BERT_3epoch_batch_8 Macro F1 = 0.15	O	0.833	0.998	0.908
	B-PER	1.0	0.0	0.0
	I-PER	1.0	0.0	0.0
	B-ORG	0.5	0.0833	0.1429
	I-ORG	1.0	0.0	0.0
	B-LOC	0.0	0.0	0.0
	I-LOC	1.0	0.0	0.0
Indic_BERT_1lakh Macro F1 = 0.14	O	0.8319	0.8516	0.8417
	B-PER	0.0	0.0	0.0
	I-PER	0.0	0.0	0.0
	B-ORG	0.0526	0.0833	0.0645
	I-ORG	0.0	0.0	0.0
	B-LOC	0.0625	0.05	0.0556
	I-LOC	0.0	0.0	0.0
Indic_BERT_50k Macro F1 = 0.14	O	0.8309	0.8559	0.8432
	B-PER	0.0	0.0	0.0
	I-PER	0.0	0.0	0.0
	B-ORG	0.0476	0.0833	0.0606
	I-ORG	0.0	0.0	0.0
	B-LOC	0.0667	0.05	0.0571
	I-LOC	0.0	0.0	0.0
filtered_Indic_BERT_1lakh Macro F1 = 0.14	O	0.9034	0.4624	0.6117
	B-PER	0.0476	0.0357	0.0408
	I-PER	0.05	0.0952	0.0656
	B-ORG	0.1154	0.25	0.1579
	I-ORG	0.0242	0.3333	0.0452
	B-LOC	0.0556	0.05	0.0526
	I-LOC	0.0	0.0	0.0
filtered_Indic_NER_1lakh Macro F1 = 0.06	O	0.9375	0.1523	0.262
	B-PER	0.0577	0.1429	0.0822
	I-PER	0.0	0.0	0.0
	B-ORG	0.0	0.0	0.0
	I-ORG	0.0263	0.2	0.0465
	B-LOC	0.0218	0.2778	0.0405
	I-LOC	0.0	0.0	0.0

Table 2: NER Model Evaluation

Model	Class	Precision	Recall	F1 Score
Indic_Ner_3epoch_batch_8 Macro F1 = 0.038	O	1.0	0.0	0.0
	B-PER	1.0	0.0	0.0
	I-PER	1.0	0.0	0.0
	B-ORG	1.0	0.0	0.0
	B-LOC	0.107	1.0	0.194
IndicBERT-3epoch_batch_16 Macro F1 = 0.11	O	0.55	1.0	0.710
	B-PER	1.0	0.0	0.0
	I-PER	1.0	0.0	0.0
	B-ORG	1.0	0.0	0.0
	I-ORG	1.0	0.0	0.0
	B-LOC	1.0	0.0	0.0
Indic_Ner_3epoch_batch_16 Macro F1 = 0.04	O	1.0	0.0	0.0
	B-PER	1.0	0.0	0.0
	I-PER	1.0	0.0	0.0
	B-ORG	1.0	0.0	0.0
	B-LOC	0.107	1.0	0.194
Indic.BERT_3epoch_batch_8 Macro F1 = 0.12	O	0.55	1.0	0.710
	B-PER	1.0	0.0	0.0
	I-PER	1.0	0.0	0.0
	B-ORG	1.0	0.0	0.0
	I-ORG	1.0	0.0	0.0
	B-LOC	1.0	0.0	0.0
Indic_BERT_1lakh Macro F1 = 0.13	O	1.0	0.0	0.0
	B-PER	1.0	0.0	0.0
	I-PER	1.0	0.0	0.0
	B-ORG	0.3	0.75	0.429
	I-ORG	0.2	1.0	0.333
	B-LOC	1.0	0.0	0.0
Indic_BERT_50k Macro F1 = 0.16	O	0.6	0.545	0.571
	B-PER	1.0	0.0	0.0
	I-PER	1.0	0.0	0.0
	B-ORG	0.3	0.75	0.429
	I-ORG	1.0	0.0	0.0
	B-LOC	1.0	0.0	0.0
filtered_Indic_BERT_1lakh Macro F1 = 0.13	O	1.0	0.0	0.0
	B-PER	1.0	0.0	0.0
	I-PER	1.0	0.0	0.0
	B-ORG	0.3	0.75	0.429
	I-ORG	0.2	1.0	0.333
	B-LOC	1.0	0.0	0.0
filtered_Indic_NER_1lakh Macro F1 = 0.00	O	1.0	0.0	0.0
	B-PER	1.0	0.0	0.0
	I-PER	1.0	0.0	0.0
	B-ORG	1.0	0.0	0.0
	B-LOC	0.0	0.0	0.0

Table 3: NER Model Evaluation

Model	Class	Precision	Recall	F1 Score
Indic_Ner_3epoch_batch_8 Macro F1 = 0.11	O	0.107	0.001	0.002
	B-PER	0.027	0.026	0.026
	I-PER	0.0	0.0	0.0
	B-ORG	0.013	0.02	0.016
	I-ORG	0.682	0.776	0.726
	B-LOC	0.003	0.077	0.006
	I-LOC	0.0	0.0	0.0
IndicBERT-3epoch_batch_16 Macro F1 = 0.13	O	0.831	0.998	0.907
	B-PER	0.0	0.0	0.0
	I-PER	0.5	0.001	0.003
	B-ORG	0.214	0.012	0.023
	I-ORG	1.0	0.0	0.0
	B-LOC	0.077	0.002	0.003
	I-LOC	1.0	0.0	0.0
Indic_Ner_3epoch_batch_16 Macro F1 = 0.14	O	0.088	0.001	0.002
	B-PER	0.037	0.037	0.037
	I-PER	0.0	0.0	0.0
	B-ORG	0.013	0.02	0.016
	I-ORG	0.692	0.778	0.732
	B-LOC	0.004	0.094	0.007
	I-LOC	0.002	0.005	0.003
Indic_BERT_3epoch_batch_8 Macro F1 = 0.13	O	0.831	0.997	0.907
	B-PER	0.071	0.001	0.003
	I-PER	0.0	0.0	0.0
	B-ORG	0.048	0.002	0.004
	I-ORG	1.0	0.0	0.0
	B-LOC	0.071	0.003	0.006
	I-LOC	0.0	0.0	0.0
Indic_BERT_1lakh Macro F1 = 0.82	O	0.973	0.977	0.975
	B-PER	0.844	0.873	0.858
	I-PER	0.917	0.892	0.905
	B-ORG	0.757	0.781	0.769
	I-ORG	0.741	0.703	0.721
	B-LOC	0.785	0.773	0.779
	I-LOC	0.701	0.543	0.612
Indic_BERT_50k Macro F1 = 0.79	O	0.969	0.974	0.972
	B-PER	0.824	0.837	0.831
	I-PER	0.91	0.875	0.892
	B-ORG	0.716	0.733	0.724
	I-ORG	0.742	0.703	0.722
	B-LOC	0.785	0.773	0.779
	I-LOC	0.701	0.543	0.612
filtered_Indic_BERT_1lakh Macro F1 = 0.48	O	0.988	0.52	0.681
	B-PER	0.69	0.684	0.687
	I-PER	0.254	0.901	0.396
	B-ORG	0.51	0.683	0.584
	I-ORG	0.086	0.862	0.156
	B-LOC	0.702	0.67	0.686
	I-LOC	0.096	0.749	0.171
filtered_Indic_NEr_1lakh Macro F1 = 0.05	O	0.844	0.077	0.141
	B-PER	0.051	0.171	0.078
	I-PER	0.0	0.0	0.0
	B-ORG	0.017	0.024	0.02
	I-ORG	0.098	0.9	0.177
	B-LOC	0.002	0.023	0.003
	I-LOC	0.0	0.0	0.0