

Experimento: EX3 - Etapa 2 (Corpus: Glossário EB + Trechos das Doutrinas)
Diário de bordo com os ajustes finos e valores dos hiperparâmetros utilizados

Resultados do treinamento (ajuste fino dos modelos pré-treinados)

Corpus: Corpus-3-GLOSSÁRIO - EB -Com uma só categoria de entidade-NER

Data: 20/11/2023

Parâmetros diferentes:

i) Notebook: IDEA-ETAPA 3-FineTuneBERT_Spacy_NER.ipynb

Vocabulário: pt_core_news_sm

Corpus: GLOSSÁRIO-EB

ii) Arquivo de configuração: config_spacy.cfg

[nlp]

Linha 12: lang = "pt"

[components.transformer.model]

Linha 52: name = "neuralmind/bert-base-portuguese-cased"

[training]

dropout = 0.2

iii) Resultados

DATA 27/11/2023

Fatiado em 1

E	#	LOSS TRANS...	LOSS NER	ENTS_F	ENTS_P	ENTS_R	SCORE
0	0	1200.22	647.50	3.24	2.22	6.02	0.03
13	200	44668.01	55378.70	78.70	73.15	85.15	0.79
26	400	1003.89	2140.67	81.74	79.70	83.89	0.82
40	600	302.29	1195.59	84.68	85.60	83.79	0.85
53	800	184.90	1019.10	84.30	85.12	83.49	0.84
66	1000	69.87	828.58	84.91	85.88	83.96	0.85
80	1200	78.23	826.50	85.73	85.86	85.60	0.86
93	1400	57.34	770.22	84.87	87.14	82.73	0.85
106	1600	51.17	716.67	85.53	86.21	84.87	0.86
119	1800	45.39	668.29	85.86	87.41	84.36	0.86
132	2000	41.30	654.55	85.48	85.35	85.60	0.85

146	2200	19.15	585.72	85.39	86.81	84.02	0.85
159	2400	34.01	554.93	86.16	86.27	86.06	0.86
172	2600	67.21	535.61	85.47	85.52	85.42	0.85
185	2800	42.32	472.92	85.21	86.07	84.36	0.85
199	3000	36.86	424.62	86.24	86.85	85.63	0.86
212	3200	19.51	375.71	85.80	85.86	85.73	0.86
225	3400	41.62	358.13	85.49	87.96	83.15	0.85
238	3600	37.31	297.40	86.01	86.48	85.55	0.86
251	3800	1.00	245.94	85.71	87.56	83.93	0.86
265	4000	91.20	259.89	85.90	85.74	86.06	0.86
278	4200	23.34	200.38	86.11	86.83	85.40	0.86
291	4400	214.75	196.66	85.95	86.85	85.07	0.86
304	4600	108.47	154.46	85.77	85.10	86.44	0.86

===== Results
=====

TOK 100.00
NER P 86.56
NER R 86.48
NER F 86.52
SPEED 1419

===== NER (per type)
=====

	P	R	F
entity	86.56	86.48	86.52

DATA 23/11/2023
Fatiado em 2

DATA 20/11/2023
Fatiado em 10

E	#	LOSS	TRANS...	LOSS NER	ENTS_F	ENTS_P	ENTS_R	SCORE
0	0	1473.74	848.98	2.69	1.96	4.31	0.03	
16	200	48096.38	64668.65	76.20	69.74	83.98	0.76	
33	400	1436.73	9784.17	82.75	83.35	82.16	0.83	
50	600	205.98	7559.55	82.88	82.63	83.14	0.83	
66	800	65.22	7053.09	83.28	83.83	82.73	0.83	
83	1000	17.06	6730.93	82.03	82.77	81.30	0.82	

100	1200	55.65	6482.07	83.29	83.93	82.65	0.83
116	1400	46.83	6099.46	84.20	84.56	83.84	0.84
133	1600	78.96	5761.79	83.03	84.70	81.42	0.83
150	1800	28.84	5369.22	83.00	85.14	80.97	0.83
166	2000	39.72	4904.78	83.71	85.41	82.08	0.84
183	2200	64.45	4477.09	83.63	85.40	81.93	0.84
200	2400	45.80	4001.17	83.56	83.31	83.81	0.84
216	2600	73.43	3478.06	81.65	81.33	81.98	0.82
233	2800	21.28	2960.24	83.72	84.29	83.15	0.84
250	3000	2.80	2461.90	84.20	85.73	82.72	0.84

===== Results
=====

TOK 100.00
NER P 84.85
NER R 84.44
NER F 84.65
SPEED 6882

===== NER (per type)
=====

	P	R	F
entity	84.85	84.44	84.65

DATA: 03/11/2023 - ARTIGO

**** neuralmind/bert-large-portuguese-cased

E	#	LOSS	TRANS...	LOSS NER	ENTS_F	ENTS_P	ENTS_R	SCORE
0	0	135.69	28.23	0.46	1.49	0.27	0.00	
200	200	2613.49	924.25	10.43	7.87	15.48	0.10	
400	400	0.00	0.00	10.42	7.76	15.85	0.10	
600	600	0.00	0.00	10.26	7.60	15.76	0.10	
800	800	0.00	0.00	10.11	7.46	15.66	0.10	
1000	1000	0.00	0.00	10.00	7.37	15.57	0.10	
1200	1200	0.00	0.00	9.89	7.27	15.48	0.10	
1400	1400	0.00	0.00	9.85	7.22	15.48	0.10	
1600	1600	0.00	0.00	9.83	7.20	15.48	0.10	
1800	1800	0.00	0.00	9.80	7.17	15.48	0.10	

TOK 100.00
NER P 9.93
NER R 17.19

NER F 12.58
SPEED 1801

P R F
entity 9.93 17.19 12.58

Corpus: Corpus-3-GLOSSÁRIO-EB-Com uma só categoria de entidade-NER
Data: 16/11/2023

Parâmetros diferentes:

i) Notebook: IDEA-ETAPA 3-FineTuneBERT_Spacy_RE.ipynb

Vocabulário: pt_core_news_sm
Corpus: glossário - eb (RE)

ii) Arquivo de configuração: \configs\config_spacy.cfg

Linha 12: lang = "pt"
Linha 19: batch_size = 500
Linha 30: name = "neuralmind/bert-base-portuguese-cased"
Linha 40: threshold = 0.5
Linha 60: max_length = 30
Linha 84: dropout = 0.2

Resultados do treinamento
23-11-2023 (fatiado em 1)

E # LOSS TRANS... LOSS RELAT... REL_MICRO_P REL_MICRO_R
REL_MICRO_F SCORE

0	0	4.22	2.74	1.67	65.23	3.26	0.03
6	100	179.73	68.07	95.35	98.49	96.90	0.97
13	200	0.01	0.46	93.18	98.49	95.76	0.96
19	300	0.02	0.49	98.49	98.49	98.49	0.98
26	400	0.02	0.38	97.55	98.49	98.02	0.98
33	500	0.03	0.29	97.59	98.35	97.97	0.98
40	600	0.01	0.23	98.58	98.52	98.55	0.99
46	700	0.00	0.10	98.41	98.73	98.57	0.99
53	800	0.00	0.03	97.78	98.68	98.23	0.98
60	900	0.00	0.01	98.09	98.68	98.38	0.98
66	1000	0.00	0.00	98.04	98.68	98.36	0.98

Random baseline:

threshold 0.00 {'rel_micro_p': '1.23', 'rel_micro_r': '100.00', 'rel_micro_f': '2.43'}

threshold 0.05	{'rel_micro_p': '1.23', 'rel_micro_r': '95.30', 'rel_micro_f': '2.43'}
threshold 0.10	{'rel_micro_p': '1.23', 'rel_micro_r': '89.89', 'rel_micro_f': '2.42'}
threshold 0.20	{'rel_micro_p': '1.22', 'rel_micro_r': '79.54', 'rel_micro_f': '2.40'}
threshold 0.30	{'rel_micro_p': '1.23', 'rel_micro_r': '70.26', 'rel_micro_f': '2.42'}
threshold 0.40	{'rel_micro_p': '1.23', 'rel_micro_r': '60.43', 'rel_micro_f': '2.42'}
threshold 0.50	{'rel_micro_p': '1.25', 'rel_micro_r': '51.15', 'rel_micro_f': '2.44'}
threshold 0.60	{'rel_micro_p': '1.24', 'rel_micro_r': '40.61', 'rel_micro_f': '2.41'}
threshold 0.70	{'rel_micro_p': '1.22', 'rel_micro_r': '29.98', 'rel_micro_f': '2.34'}
threshold 0.80	{'rel_micro_p': '1.22', 'rel_micro_r': '20.02', 'rel_micro_f': '2.30'}
threshold 0.90	{'rel_micro_p': '1.24', 'rel_micro_r': '10.23', 'rel_micro_f': '2.22'}
threshold 0.99	{'rel_micro_p': '1.07', 'rel_micro_r': '0.88', 'rel_micro_f': '0.96'}
threshold 1.00	{'rel_micro_p': '0.00', 'rel_micro_r': '0.00', 'rel_micro_f': '0.00'}

Results of the trained model:

threshold 0.00	{'rel_micro_p': '1.23', 'rel_micro_r': '100.00', 'rel_micro_f': '2.43'}
threshold 0.05	{'rel_micro_p': '97.13', 'rel_micro_r': '98.41', 'rel_micro_f': '97.77'}
threshold 0.10	{'rel_micro_p': '97.51', 'rel_micro_r': '98.37', 'rel_micro_f': '97.94'}
threshold 0.20	{'rel_micro_p': '97.78', 'rel_micro_r': '98.37', 'rel_micro_f': '98.08'}
threshold 0.30	{'rel_micro_p': '97.90', 'rel_micro_r': '98.37', 'rel_micro_f': '98.13'}
threshold 0.40	{'rel_micro_p': '97.98', 'rel_micro_r': '98.37', 'rel_micro_f': '98.17'}
threshold 0.50	{'rel_micro_p': '98.06', 'rel_micro_r': '98.37', 'rel_micro_f': '98.21'}
threshold 0.60	{'rel_micro_p': '98.05', 'rel_micro_r': '98.33', 'rel_micro_f': '98.19'}
threshold 0.70	{'rel_micro_p': '98.05', 'rel_micro_r': '98.33', 'rel_micro_f': '98.19'}
threshold 0.80	{'rel_micro_p': '98.05', 'rel_micro_r': '98.33', 'rel_micro_f': '98.19'}
threshold 0.90	{'rel_micro_p': '98.09', 'rel_micro_r': '98.17', 'rel_micro_f': '98.13'}
threshold 0.99	{'rel_micro_p': '98.76', 'rel_micro_r': '95.30', 'rel_micro_f': '97.00'}
threshold 1.00	{'rel_micro_p': '99.38', 'rel_micro_r': '63.85', 'rel_micro_f': '77.75'}

 Resultados do treinamento
 23-11-2023 (fatiado em 2)

E	#	LOSS TRANS...	LOSS RELAT...	REL_MICRO_P	REL_MICRO_R	REL_MICRO_F	SCORE
---	---	---------------	---------------	-------------	-------------	-------------	-------

0	0	5.91	2.94	0.46	62.55	0.91	0.01
8	100	168.83	66.49	48.16	13.26	20.80	0.21
16	200	0.11	2.74	50.80	19.80	28.50	0.28
25	300	0.11	2.63	41.69	35.65	38.43	0.38
33	400	0.14	2.62	76.11	3.23	6.20	0.06
41	500	0.13	2.76	52.50	30.32	38.44	0.38
50	600	0.07	2.36	61.26	17.07	26.70	0.27
58	700	0.03	1.99	52.93	35.77	42.69	0.43
66	800	0.01	1.83	52.80	31.79	39.68	0.40
75	900	0.00	1.76	52.37	35.66	42.43	0.42
83	1000	0.01	1.73	49.11	41.93	45.24	0.45

Random baseline:

threshold 0.00	{'rel_micro_p': '0.34', 'rel_micro_r': '100.00', 'rel_micro_f': '0.68'}
threshold 0.05	{'rel_micro_p': '0.34', 'rel_micro_r': '94.81', 'rel_micro_f': '0.68'}
threshold 0.10	{'rel_micro_p': '0.34', 'rel_micro_r': '89.86', 'rel_micro_f': '0.68'}
threshold 0.20	{'rel_micro_p': '0.34', 'rel_micro_r': '80.21', 'rel_micro_f': '0.68'}
threshold 0.30	{'rel_micro_p': '0.34', 'rel_micro_r': '70.72', 'rel_micro_f': '0.69'}
threshold 0.40	{'rel_micro_p': '0.35', 'rel_micro_r': '60.78', 'rel_micro_f': '0.69'}
threshold 0.50	{'rel_micro_p': '0.35', 'rel_micro_r': '50.93', 'rel_micro_f': '0.69'}
threshold 0.60	{'rel_micro_p': '0.35', 'rel_micro_r': '41.03', 'rel_micro_f': '0.69'}
threshold 0.70	{'rel_micro_p': '0.35', 'rel_micro_r': '30.89', 'rel_micro_f': '0.70'}
threshold 0.80	{'rel_micro_p': '0.36', 'rel_micro_r': '21.00', 'rel_micro_f': '0.70'}
threshold 0.90	{'rel_micro_p': '0.39', 'rel_micro_r': '11.50', 'rel_micro_f': '0.76'}
threshold 0.99	{'rel_micro_p': '0.34', 'rel_micro_r': '1.01', 'rel_micro_f': '0.51'}
threshold 1.00	{'rel_micro_p': '0.28', 'rel_micro_r': '0.08', 'rel_micro_f': '0.12'}

Results of the trained model:

threshold 0.00	{'rel_micro_p': '0.34', 'rel_micro_r': '100.00', 'rel_micro_f': '0.68'}
threshold 0.05	{'rel_micro_p': '42.96', 'rel_micro_r': '57.20', 'rel_micro_f': '49.07'}
threshold 0.10	{'rel_micro_p': '45.03', 'rel_micro_r': '55.19', 'rel_micro_f': '49.59'}
threshold 0.20	{'rel_micro_p': '47.28', 'rel_micro_r': '52.09', 'rel_micro_f': '49.57'}
threshold 0.30	{'rel_micro_p': '49.18', 'rel_micro_r': '50.36', 'rel_micro_f': '49.76'}
threshold 0.40	{'rel_micro_p': '50.63', 'rel_micro_r': '48.55', 'rel_micro_f': '49.57'}
threshold 0.50	{'rel_micro_p': '51.98', 'rel_micro_r': '46.98', 'rel_micro_f': '49.36'}
threshold 0.60	{'rel_micro_p': '53.49', 'rel_micro_r': '44.65', 'rel_micro_f': '48.67'}
threshold 0.70	{'rel_micro_p': '54.75', 'rel_micro_r': '41.51', 'rel_micro_f': '47.22'}
threshold 0.80	{'rel_micro_p': '56.24', 'rel_micro_r': '37.69', 'rel_micro_f': '45.13'}
threshold 0.90	{'rel_micro_p': '56.90', 'rel_micro_r': '30.85', 'rel_micro_f': '40.01'}
threshold 0.99	{'rel_micro_p': '50.32', 'rel_micro_r': '9.41', 'rel_micro_f': '15.86'}
threshold 1.00	{'rel_micro_p': '29.69', 'rel_micro_r': '1.53', 'rel_micro_f': '2.91'}

Resultados do treinamento
20-11-2023 (fatiado em 10)

E	#	LOSS TRANS...	LOSS RELAT...	REL_MICRO_P	REL_MICRO_R	REL_MICRO_F	SCORE
---	---	---------------	---------------	-------------	-------------	-------------	-------

0	0	2.93	2.64	0.44	60.12	0.87	0.01
25	100	140.68	63.07	40.33	13.48	20.20	0.20
50	200	0.06	2.37	31.98	49.30	38.80	0.39
75	300	0.04	1.99	37.75	31.65	34.43	0.34
100	400	0.01	1.66	36.19	43.35	39.45	0.39
125	500	0.00	1.62	38.05	31.68	34.58	0.35
150	600	0.00	1.60	38.97	29.87	33.82	0.34
175	700	0.00	1.59	38.30	29.80	33.52	0.34
200	800	0.00	1.59	40.42	24.70	30.67	0.31
225	900	0.00	1.58	37.82	34.16	35.89	0.36
250	1000	0.00	1.57	36.99	32.64	34.68	0.35

Random baseline:

threshold 0.00	{'rel_micro_p': '0.34', 'rel_micro_r': '100.00', 'rel_micro_f': '0.67'}
threshold 0.05	{'rel_micro_p': '0.34', 'rel_micro_r': '95.39', 'rel_micro_f': '0.67'}
threshold 0.10	{'rel_micro_p': '0.34', 'rel_micro_r': '91.02', 'rel_micro_f': '0.68'}
threshold 0.20	{'rel_micro_p': '0.33', 'rel_micro_r': '78.88', 'rel_micro_f': '0.66'}
threshold 0.30	{'rel_micro_p': '0.34', 'rel_micro_r': '69.90', 'rel_micro_f': '0.67'}
threshold 0.40	{'rel_micro_p': '0.34', 'rel_micro_r': '60.92', 'rel_micro_f': '0.68'}
threshold 0.50	{'rel_micro_p': '0.34', 'rel_micro_r': '50.00', 'rel_micro_f': '0.67'}
threshold 0.60	{'rel_micro_p': '0.34', 'rel_micro_r': '40.66', 'rel_micro_f': '0.68'}
threshold 0.70	{'rel_micro_p': '0.35', 'rel_micro_r': '31.31', 'rel_micro_f': '0.70'}
threshold 0.80	{'rel_micro_p': '0.34', 'rel_micro_r': '20.15', 'rel_micro_f': '0.67'}
threshold 0.90	{'rel_micro_p': '0.34', 'rel_micro_r': '9.95', 'rel_micro_f': '0.65'}
threshold 0.99	{'rel_micro_p': '0.25', 'rel_micro_r': '0.73', 'rel_micro_f': '0.37'}
threshold 1.00	{'rel_micro_p': '0.42', 'rel_micro_r': '0.12', 'rel_micro_f': '0.19'}

Results of the trained model:

threshold 0.00	{'rel_micro_p': '0.34', 'rel_micro_r': '100.00', 'rel_micro_f': '0.67'}
threshold 0.05	{'rel_micro_p': '30.69', 'rel_micro_r': '68.93', 'rel_micro_f': '42.47'}
threshold 0.10	{'rel_micro_p': '32.80', 'rel_micro_r': '64.81', 'rel_micro_f': '43.56'}
threshold 0.20	{'rel_micro_p': '34.52', 'rel_micro_r': '57.89', 'rel_micro_f': '43.25'}
threshold 0.30	{'rel_micro_p': '37.04', 'rel_micro_r': '54.61', 'rel_micro_f': '44.14'}
threshold 0.40	{'rel_micro_p': '37.74', 'rel_micro_r': '49.15', 'rel_micro_f': '42.70'}
threshold 0.50	{'rel_micro_p': '38.45', 'rel_micro_r': '42.84', 'rel_micro_f': '40.53'}
threshold 0.60	{'rel_micro_p': '38.78', 'rel_micro_r': '36.89', 'rel_micro_f': '37.81'}
threshold 0.70	{'rel_micro_p': '39.53', 'rel_micro_r': '32.52', 'rel_micro_f': '35.69'}
threshold 0.80	{'rel_micro_p': '39.64', 'rel_micro_r': '26.46', 'rel_micro_f': '31.73'}
threshold 0.90	{'rel_micro_p': '43.80', 'rel_micro_r': '19.30', 'rel_micro_f': '26.79'}
threshold 0.99	{'rel_micro_p': '35.48', 'rel_micro_r': '1.33', 'rel_micro_f': '2.57'}
threshold 1.00	{'rel_micro_p': '0.00', 'rel_micro_r': '0.00', 'rel_micro_f': '0.00'}

03-11-2023 - ARTIGO

RELATION EXTRECTION

E	#	LOSS TRANS...	LOSS RELAT...	REL_MICRO_P	REL_MICRO_R	REL_MICRO_F	SCORE
---	---	---------------	---------------	-------------	-------------	-------------	-------

0	0	4.21	2.77	0.42	65.78	0.83	0.01
100	100	88.17	62.32	0.00	0.00	0.00	0.00
200	200	0.01	2.44	0.00	0.00	0.00	0.00
300	300	0.01	1.64	0.00	0.00	0.00	0.00
400	400	0.00	1.25	0.00	0.00	0.00	0.00
500	500	0.00	1.13	0.00	0.00	0.00	0.00
600	600	0.00	1.10	0.00	0.00	0.00	0.00
700	700	0.00	1.09	0.00	0.00	0.00	0.00
800	800	0.00	1.07	0.00	0.00	0.00	0.00
900	900	0.00	1.07	0.00	0.00	0.00	0.00
1000	1000	0.00	1.06	0.00	0.00	0.00	0.00

✓ Saved pipeline to output directory
training/model-last

Random baseline:

threshold 0.00	{'rel_micro_p': '0.29', 'rel_micro_r': '100.00', 'rel_micro_f': '0.58'}
threshold 0.05	{'rel_micro_p': '0.29', 'rel_micro_r': '94.90', 'rel_micro_f': '0.58'}
threshold 0.10	{'rel_micro_p': '0.29', 'rel_micro_r': '89.72', 'rel_micro_f': '0.58'}
threshold 0.20	{'rel_micro_p': '0.29', 'rel_micro_r': '79.87', 'rel_micro_f': '0.58'}
threshold 0.30	{'rel_micro_p': '0.29', 'rel_micro_r': '70.53', 'rel_micro_f': '0.58'}
threshold 0.40	{'rel_micro_p': '0.29', 'rel_micro_r': '60.02', 'rel_micro_f': '0.58'}
threshold 0.50	{'rel_micro_p': '0.29', 'rel_micro_r': '50.48', 'rel_micro_f': '0.58'}
threshold 0.60	{'rel_micro_p': '0.29', 'rel_micro_r': '39.92', 'rel_micro_f': '0.58'}
threshold 0.70	{'rel_micro_p': '0.29', 'rel_micro_r': '30.10', 'rel_micro_f': '0.58'}
threshold 0.80	{'rel_micro_p': '0.29', 'rel_micro_r': '19.90', 'rel_micro_f': '0.57'}
threshold 0.90	{'rel_micro_p': '0.29', 'rel_micro_r': '10.08', 'rel_micro_f': '0.57'}
threshold 0.99	{'rel_micro_p': '0.35', 'rel_micro_r': '1.22', 'rel_micro_f': '0.54'}
threshold 1.00	{'rel_micro_p': '0.24', 'rel_micro_r': '0.08', 'rel_micro_f': '0.13'}

Results of the trained model:

threshold 0.00	{'rel_micro_p': '0.29', 'rel_micro_r': '100.00', 'rel_micro_f': '0.58'}
threshold 0.05	{'rel_micro_p': '0.29', 'rel_micro_r': '100.00', 'rel_micro_f': '0.58'}
threshold 0.10	{'rel_micro_p': '0.29', 'rel_micro_r': '100.00', 'rel_micro_f': '0.58'}
threshold 0.20	{'rel_micro_p': '0.29', 'rel_micro_r': '100.00', 'rel_micro_f': '0.58'}
threshold 0.30	{'rel_micro_p': '0.31', 'rel_micro_r': '98.84', 'rel_micro_f': '0.63'}
threshold 0.40	{'rel_micro_p': '0.35', 'rel_micro_r': '88.17', 'rel_micro_f': '0.70'}
threshold 0.50	{'rel_micro_p': '0.36', 'rel_micro_r': '56.48', 'rel_micro_f': '0.72'}
threshold 0.60	{'rel_micro_p': '0.28', 'rel_micro_r': '19.42', 'rel_micro_f': '0.56'}
threshold 0.70	{'rel_micro_p': '0.14', 'rel_micro_r': '2.46', 'rel_micro_f': '0.27'}
threshold 0.80	{'rel_micro_p': '0.06', 'rel_micro_r': '0.08', 'rel_micro_f': '0.07'}
threshold 0.90	{'rel_micro_p': '0.00', 'rel_micro_r': '0.00', 'rel_micro_f': '0.00'}
threshold 0.99	{'rel_micro_p': '0.00', 'rel_micro_r': '0.00', 'rel_micro_f': '0.00'}
threshold 1.00	{'rel_micro_p': '0.00', 'rel_micro_r': '0.00', 'rel_micro_f': '0.00'}

Rodadas de treinamento - IDEA-C2

Curadoria: até Área de sobrevoos restritos

Data: 10/05/2024

1.1) TAREFA: NER

CORPUS:

PIPELINE: pt_core_news_sm

BATCH_SIZE = 128 (Linha: 14)

MAX_BATCH_ITEMS = 4096 (Linha: 48)

NAME = "neuralmind/bert-base-portuguese-cased" (Linha: 53)

DROPOUT = 0.2 (Linha: 92)

MAX_STEPS = 20000 (Linha: 95)

E	#	LOSS TRANS...	LOSS NER	ENTS_F	ENTS_P	ENTS_R	SCORE
0	0	1671.67	939.52	3.48	2.41	6.24	0.03
9	200	49222.03	60791.90	73.45	69.88	77.42	0.73
18	400	2677.58	4832.45	79.57	82.34	76.98	0.80
27	600	605.18	1888.64	79.96	80.74	79.20	0.80
36	800	413.13	1594.92	80.52	78.77	82.36	0.81
46	1000	199.35	1348.41	81.25	83.70	78.94	0.81
55	1200	151.62	1248.63	81.31	80.98	81.64	0.81
64	1400	142.18	1153.32	82.03	83.22	80.87	0.82
74	1600	117.37	1084.00	81.51	84.19	79.00	0.82
83	1800	119.75	1029.68	81.53	83.16	79.97	0.82
92	2000	30.51	876.74	81.81	84.74	79.08	0.82
101	2200	85.16	832.33	82.12	82.69	81.55	0.82
110	2400	79.07	808.45	82.56	85.24	80.05	0.83
120	2600	39.69	693.82	81.98	82.99	80.99	0.82
129	2800	99.84	641.33	81.75	81.84	81.66	0.82
138	3000	108.57	625.90	81.86	82.08	81.63	0.82
147	3200	78.23	514.53	82.55	85.47	79.82	0.83
156	3400	157.04	516.13	82.74	84.49	81.06	0.83
166	3600	68.87	426.76	82.00	83.48	80.58	0.82
175	3800	43.80	333.14	82.85	83.85	81.88	0.83
184	4000	67.09	289.85	82.68	85.37	80.15	0.83
193	4200	113.32	272.91	82.16	82.55	81.78	0.82
202	4400	143.05	240.81	82.16	82.60	81.71	0.82
211	4600	47.96	169.98	82.49	85.44	79.75	0.82
221	4800	99.71	167.22	82.90	83.83	81.99	0.83
230	5000	150.54	163.25	83.59	85.59	81.69	0.84
239	5200	91.87	128.42	82.51	85.10	80.07	0.83
248	5400	59.02	92.47	82.55	83.05	82.07	0.83
258	5600	51.95	77.76	83.09	85.36	80.94	0.83
267	5800	38.70	61.79	83.29	85.29	81.39	0.83
276	6000	74.66	45.73	83.30	85.09	81.59	0.83
285	6200	25.31	33.14	83.38	85.28	81.55	0.83
295	6400	52.15	31.80	82.46	83.07	81.87	0.82
304	6600	42.80	40.97	83.18	86.56	80.05	0.83

✓ Saved pipeline to output directory

/content/drive/MyDrive/Doutorado/IDEA/NER/outputs/model-last

===== Training stats

=====

Language: pt

Training pipeline: transformer, ner

693 training docs

1618 evaluation docs

- ✓ No overlap between training and evaluation data
- ⚠ Low number of examples to train a new pipeline (693)

===== Vocab & Vectors
=====

i 31342 total word(s) in the data (5453 unique)
i No word vectors present in the package

===== Named Entity Recognition =====

i 1 label(s)
0 missing value(s) (tokens with '-' label)
✓ Good amount of examples for all labels
✓ Examples without occurrences available for all labels
✓ No entities consisting of or starting/ending with whitespace
✓ No entities crossing sentence boundaries

===== Summary =====

✓ 7 checks passed
⚠ 1 warning

TOK 100.00
NER P 86.08
NER R 82.47
NER F 84.24
SPEED 4608

TAREFA: Relation Extraction

batch_size = 2000 (Linha: 19)
name = "neuralmind/bert-base-portuguese-cased" (Linha: 30)
threshold = 0.3 (Linha: 40)
max_length = 250 (Linha: 60)
max_steps = 1000 (Linha: 88)
size = 2000 (Linha: 99)
buffer = 256 (Linha: 100)

===== Training pipeline
=====

i Pipeline: ['transformer', 'relation_extractor']
i Initial learn rate: 0.0
E # LOSS TRANS... LOSS RELAT... REL_MICRO_P REL_MICRO_R
REL_MICRO_F SCORE

--- -----

0	0	3.95	3.08	1.12	98.53	2.22	0.02
4	100	180.71	71.51	88.35	87.51	87.93	0.88
9	200	0.08	2.26	93.23	85.35	89.12	0.89
13	300	0.07	2.05	86.40	88.77	87.57	0.88
18	400	0.07	1.53	89.53	88.62	89.07	0.89
23	500	0.04	1.13	89.61	89.04	89.33	0.89
27	600	0.02	0.96	90.43	88.84	89.63	0.90
32	700	0.01	0.79	90.79	88.64	89.70	0.90
36	800	0.00	0.78	91.65	88.70	90.15	0.90
41	900	0.00	0.75	91.02	89.10	90.05	0.90
46	1000	0.00	0.77	90.36	89.04	89.70	0.90

✓ Saved pipeline to output directory

training/model-last

Random baseline:

threshold 0.00	{'rel_micro_p': '1.09', 'rel_micro_r': '100.00', 'rel_micro_f': '2.15'}
threshold 0.05	{'rel_micro_p': '1.09', 'rel_micro_r': '95.50', 'rel_micro_f': '2.16'}
threshold 0.10	{'rel_micro_p': '1.09', 'rel_micro_r': '90.59', 'rel_micro_f': '2.16'}
threshold 0.20	{'rel_micro_p': '1.09', 'rel_micro_r': '80.53', 'rel_micro_f': '2.15'}
threshold 0.30	{'rel_micro_p': '1.09', 'rel_micro_r': '70.44', 'rel_micro_f': '2.15'}
threshold 0.40	{'rel_micro_p': '1.08', 'rel_micro_r': '59.93', 'rel_micro_f': '2.13'}
threshold 0.50	{'rel_micro_p': '1.08', 'rel_micro_r': '49.91', 'rel_micro_f': '2.12'}
threshold 0.60	{'rel_micro_p': '1.09', 'rel_micro_r': '40.31', 'rel_micro_f': '2.13'}
threshold 0.70	{'rel_micro_p': '1.10', 'rel_micro_r': '30.37', 'rel_micro_f': '2.12'}
threshold 0.80	{'rel_micro_p': '1.11', 'rel_micro_r': '20.54', 'rel_micro_f': '2.11'}
threshold 0.90	{'rel_micro_p': '1.12', 'rel_micro_r': '10.36', 'rel_micro_f': '2.03'}
threshold 0.99	{'rel_micro_p': '1.18', 'rel_micro_r': '1.09', 'rel_micro_f': '1.14'}
threshold 1.00	{'rel_micro_p': '1.66', 'rel_micro_r': '0.15', 'rel_micro_f': '0.27'}

Results of the trained model:

threshold 0.00	{'rel_micro_p': '1.09', 'rel_micro_r': '100.00', 'rel_micro_f': '2.15'}
threshold 0.05	{'rel_micro_p': '85.53', 'rel_micro_r': '91.08', 'rel_micro_f': '88.21'}
threshold 0.10	{'rel_micro_p': '89.72', 'rel_micro_r': '90.57', 'rel_micro_f': '90.14'}
threshold 0.20	{'rel_micro_p': '92.07', 'rel_micro_r': '90.06', 'rel_micro_f': '91.05'}
threshold 0.30	{'rel_micro_p': '92.97', 'rel_micro_r': '89.69', 'rel_micro_f': '91.30'}
threshold 0.40	{'rel_micro_p': '93.58', 'rel_micro_r': '89.33', 'rel_micro_f': '91.40'}
threshold 0.50	{'rel_micro_p': '94.06', 'rel_micro_r': '88.99', 'rel_micro_f': '91.45'}
threshold 0.60	{'rel_micro_p': '94.64', 'rel_micro_r': '88.45', 'rel_micro_f': '91.44'}
threshold 0.70	{'rel_micro_p': '95.21', 'rel_micro_r': '87.89', 'rel_micro_f': '91.40'}
threshold 0.80	{'rel_micro_p': '95.95', 'rel_micro_r': '87.07', 'rel_micro_f': '91.29'}
threshold 0.90	{'rel_micro_p': '96.61', 'rel_micro_r': '85.92', 'rel_micro_f': '90.95'}
threshold 0.99	{'rel_micro_p': '98.12', 'rel_micro_r': '81.06', 'rel_micro_f': '88.78'}
threshold 1.00	{'rel_micro_p': '98.53', 'rel_micro_r': '66.96', 'rel_micro_f': '79.73'}

Teste:

Texto submetido:

text =["As Operações (OP) ÁGATA são coordenadas pelo "]
text = text + ["Ministério da Defesa (MD), por meio da Marinha (MB), do Exército (EB) e da Aeronáutica (FAB), "]
text = text + ["com a participação de órgãos de segurança e de fiscalização pública federais, estaduais e municipais (OSP). "]
text = text + ["Destinam-se ao combate a crimes transfronteiriços, com a intensificação da presença do Estado nas áreas de "]
text = text + ["fronteira e a integração com outros órgãos e organizações da sociedade civil. "]
text = text + ["As Forças Armadas (FA) auxiliam a segurança pública (SegPb) no estabelecimento de postos de bloqueio, "]
text = text + ["no controle de rios e estradas, em revistas de pessoas, embarcações, veículos e aeronaves e "]
text = text + ["na execução de patrulhas fluviais, terrestres e aéreas. "]
text = text + ["No ano de 2022, foram realizadas mais de 610 operações singulares (OpSing) Ágata "]
text = text + ["(quando somente uma das Forças Armadas participa) e 04 operações conjuntas (OpCj) Ágata e "]
text = text + ["foram executadas ações (AC) de 3.729 patrulhas, mais de "]
text = text + ["140 mil revistas e vistorias, 429 inspeções de aeronaves e 433 interceptações de aeronaves. "]
text = text + ["APROVEITAMENTO DO ÊXITO – Tipo de operação ofensiva que se segue a um ataque bem sucedido e que, normalmente, tem início quando a força inimiga encontra-se em dificuldades para manter suas posições, visando anular sua # capacidade de reorganizar-se ou de realizar um movimento retrógrado."]

RESPOSTA:

```
[[0, 'OP', 'entity'], [1, 'Defesa', 'entity'], [2, 'segurança', 'entity'], [3, 'fiscalização pública', 'entity'], [4, 'OSP', 'entity'], [5, 'combate', 'entity'], [6, 'presença', 'entity'], [7, 'Estado', 'entity'], [8, 'integração', 'entity'], [9, 'órgãos', 'entity'], [10, 'Forças Armadas', 'entity'], [11, 'segurança pública', 'entity'], [12, 'bloqueio', 'entity'], [13, 'controle', 'entity'], [15, 'operações conjuntas', 'entity'], [16, 'APROVEITAMENTO DO ÊXITO', 'entity'], [17, 'operação ofensiva', 'entity'], [18, 'ataque', 'entity'], [19, 'força', 'entity'], [20, 'movimento', 'entity']]
```

<built-in method count of list object at 0x7a2cb3a11100>

```
[[0, 'APROVEITAMENTO DO ÊXITO', 'associated_with', 'ataque'], [1, 'APROVEITAMENTO DO ÊXITO', 'associated_with', 'força'], [2, 'APROVEITAMENTO DO ÊXITO', 'associated_with', 'movimento'], [3, 'operação ofensiva', 'type_of', 'APROVEITAMENTO DO ÊXITO']]
```

Data: 12/05/2024

TAREFA: NER

CORPUS:

PIPELINE: pt_core_news_md

BATCH_SIZE = 128 (Linha: 14)

MAX_BATCH_ITEMS = 4096 (Linha: 48)

NAME = "neuralmind/bert-base-portuguese-cased" (Linha: 53)

DROPOUT = 0.2 (Linha: 92)

MAX_STEPS = 20000 (Linha: 95)

===== Training pipeline =====

i Pipeline: ['transformer', 'ner']

i Initial learn rate: 0.0

E	#	LOSS TRANS...	LOSS NER	ENTS_F	ENTS_P	ENTS_R	SCORE
---	---	---------------	----------	--------	--------	--------	-------

0	0	945.80	456.82	3.47	2.41	6.20	0.03
9	200	47140.07	56578.23	76.70	80.82	72.98	0.77
19	400	2262.57	3931.94	80.25	81.70	78.85	0.80
29	600	462.21	1393.66	80.56	82.47	78.74	0.81
39	800	308.99	1150.62	80.81	80.74	80.87	0.81
49	1000	182.06	967.06	81.38	82.12	80.65	0.81
59	1200	170.94	924.33	81.57	84.80	78.57	0.82
70	1400	140.03	845.75	82.39	84.73	80.18	0.82
80	1600	60.62	740.08	82.95	84.70	81.26	0.83
90	1800	86.81	723.72	82.21	85.52	79.15	0.82
99	2000	91.45	696.95	82.54	83.85	81.27	0.83
109	2200	29.84	613.42	81.45	81.18	81.73	0.81
119	2400	47.94	572.95	81.84	81.61	82.07	0.82
129	2600	40.13	543.33	81.55	83.27	79.89	0.82
139	2800	58.57	503.59	82.91	85.30	80.64	0.83
149	3000	67.71	488.83	82.74	83.67	81.83	0.83
159	3200	34.50	419.29	82.64	85.09	80.32	0.83

✓ Saved pipeline to output directory

/content/drive/MyDrive/Doutorado/IDEA/NER/outputs/model-last

===== Results =====

TOK 100.00
NER P 84.42
NER R 79.07
NER F 81.66
SPEED 5133

===== NER (per type) =====

	P	R	F
entity	84.42	79.07	81.66

TAREFA: Relation Extraction

batch_size = 2000 (Linha: 19)

name = "neuralmind/bert-base-portuguese-cased" (Linha: 30)

threshold = 0.3 (Linha: 40)

max_length = 250 (Linha: 60)

max_steps = 1000 (Linha: 88)

size = 2000 (Linha: 99)

buffer = 256 (Linha: 100)

i Pipeline: ['transformer', 'relation_extractor']

i Initial learn rate: 0.0

E # LOSS TRANS... LOSS RELAT... REL_MICRO_P REL_MICRO_R
REL_MICRO_F SCORE

0	0	0.75	2.85	1.12	98.55	2.21	0.02
4	100	172.46	70.56	89.63	88.12	88.87	0.89
10	200	0.13	3.00	88.99	88.03	88.51	0.89
15	300	0.09	2.29	92.87	87.95	90.34	0.90
20	400	0.09	1.63	92.18	89.06	90.59	0.91
25	500	0.05	1.13	92.50	87.62	90.00	0.90
30	600	0.04	0.93	91.84	89.51	90.66	0.91
35	700	0.02	0.72	91.04	89.72	90.37	0.90
40	800	0.03	0.62	94.04	88.44	91.15	0.91
45	900	0.01	0.56	92.74	89.24	90.95	0.91
49	1000	0.00	0.54	92.79	89.21	90.96	0.91

✓ Saved pipeline to output directory
training/model-last

Random baseline:

threshold 0.00	{'rel_micro_p': '0.71', 'rel_micro_r': '100.00', 'rel_micro_f': '1.41'}
threshold 0.05	{'rel_micro_p': '0.71', 'rel_micro_r': '94.86', 'rel_micro_f': '1.41'}
threshold 0.10	{'rel_micro_p': '0.71', 'rel_micro_r': '90.16', 'rel_micro_f': '1.42'}
threshold 0.20	{'rel_micro_p': '0.71', 'rel_micro_r': '79.92', 'rel_micro_f': '1.41'}
threshold 0.30	{'rel_micro_p': '0.70', 'rel_micro_r': '69.07', 'rel_micro_f': '1.39'}
threshold 0.40	{'rel_micro_p': '0.70', 'rel_micro_r': '59.10', 'rel_micro_f': '1.39'}
threshold 0.50	{'rel_micro_p': '0.70', 'rel_micro_r': '48.99', 'rel_micro_f': '1.38'}
threshold 0.60	{'rel_micro_p': '0.69', 'rel_micro_r': '38.86', 'rel_micro_f': '1.36'}
threshold 0.70	{'rel_micro_p': '0.68', 'rel_micro_r': '28.40', 'rel_micro_f': '1.32'}
threshold 0.80	{'rel_micro_p': '0.67', 'rel_micro_r': '18.83', 'rel_micro_f': '1.30'}
threshold 0.90	{'rel_micro_p': '0.69', 'rel_micro_r': '9.59', 'rel_micro_f': '1.28'}
threshold 0.99	{'rel_micro_p': '0.81', 'rel_micro_r': '1.12', 'rel_micro_f': '0.94'}
threshold 1.00	{'rel_micro_p': '0.66', 'rel_micro_r': '0.09', 'rel_micro_f': '0.16'}

Results of the trained model:

threshold 0.00	{'rel_micro_p': '0.71', 'rel_micro_r': '100.00', 'rel_micro_f': '1.41'}
threshold 0.05	{'rel_micro_p': '77.77', 'rel_micro_r': '84.86', 'rel_micro_f': '81.16'}
threshold 0.10	{'rel_micro_p': '84.54', 'rel_micro_r': '84.10', 'rel_micro_f': '84.32'}
threshold 0.20	{'rel_micro_p': '88.06', 'rel_micro_r': '83.09', 'rel_micro_f': '85.50'}
threshold 0.30	{'rel_micro_p': '89.74', 'rel_micro_r': '82.36', 'rel_micro_f': '85.89'}
threshold 0.40	{'rel_micro_p': '91.02', 'rel_micro_r': '81.82', 'rel_micro_f': '86.18'}
threshold 0.50	{'rel_micro_p': '91.93', 'rel_micro_r': '81.28', 'rel_micro_f': '86.28'}
threshold 0.60	{'rel_micro_p': '92.75', 'rel_micro_r': '80.68', 'rel_micro_f': '86.30'}
threshold 0.70	{'rel_micro_p': '93.52', 'rel_micro_r': '80.08', 'rel_micro_f': '86.28'}
threshold 0.80	{'rel_micro_p': '94.41', 'rel_micro_r': '78.94', 'rel_micro_f': '85.98'}
threshold 0.90	{'rel_micro_p': '95.55', 'rel_micro_r': '76.79', 'rel_micro_f': '85.15'}
threshold 0.99	{'rel_micro_p': '98.08', 'rel_micro_r': '61.67', 'rel_micro_f': '75.73'}
threshold 1.00	{'rel_micro_p': '97.17', 'rel_micro_r': '5.37', 'rel_micro_f': '10.17'}

TESTE:

```

text =["As Operações (OP) ÁGATA são coordenadas pelo "]
text = text + ["Ministério da Defesa (MD), por meio da Marinha (MB), do Exército (EB) e da
Aeronáutica (FAB), "]
text = text + ["com a participação de órgãos de segurança e de fiscalização pública federais,
estaduais e municipais (OSP). "]
text = text + ["Destinam-se ao combate a crimes transfronteiriços, com a intensificação da
presença do Estado nas áreas de "]
text = text + ["fronteira e a integração com outros órgãos e organizações da sociedade civil.
"]
text = text + ["As Forças Armadas (FA) auxiliam a segurança pública (SegPb) no
estabelecimento de postos de bloqueio, "]
text = text + ["no controle de rios e estradas, em revistas de pessoas, embarcações,
veículos e aeronaves e "]
text = text + ["na execução de patrulhas fluviais, terrestres e aéreas. "]
text = text + ["No ano de 2022, foram realizadas mais de 610 operações singulares
(OpSing) Ágata "]
text = text + ["(quando somente uma das Forças Armadas participa) e 04 operações
conjuntas (OpCj) Ágata e "]
text = text + ["foram executadas ações (AC) de 3.729 patrulhas, mais de "]
text = text + ["140 mil revistas e vistorias, 429 inspeções de aeronaves e 433 interceptações
de aeronaves. "]
text = text + ["APROVEITAMENTO DO ÊXITO – Tipo de operação ofensiva que se segue a
um ataque bem sucedido e que, normalmente, tem início quando a força inimiga
encontra-se em dificuldades para manter suas posições, visando anular sua
# capacidade de reorganizar-se ou de realizar um movimento retrógrado.""]

```

RESULTADOS:

SEM QUEBRA DE SENTENÇA: NAO

```

[[0, 'Operações (OP)', 'entity'], [1, 'ÁGATA', 'entity'], [2, 'Defesa', 'entity'], [3, 'segurança',
'entity'], [4, 'combate', 'entity'], [5, 'presença', 'entity'], [6, 'Estado', 'entity'], [7, 'fronteira',
'entity'], [8, 'integração', 'entity'], [9, 'Forças Armadas', 'entity'], [10, 'segurança pública',
'entity'], [11, 'bloqueio', 'entity'], [12, 'controle', 'entity'], [13, 'Ágata', 'entity'], [16,
'APROVEITAMENTO DO ÊXITO', 'entity'], [17, 'operação ofensiva', 'entity'], [18, 'ataque',
'entity'], [19, 'força', 'entity'], [20, 'movimento', 'entity']]
<built-in method count of list object at 0x7de41dd8ff80>
[[0, 'APROVEITAMENTO DO ÊXITO', 'associated_with', 'ataque'], [1, 'APROVEITAMENTO
DO ÊXITO', 'associated_with', 'força'], [2, 'APROVEITAMENTO DO ÊXITO',
'associated_with', 'movimento'], [3, 'operação ofensiva', 'type_of', 'APROVEITAMENTO DO
ÊXITO']]

```

SEM QUEBRA DE SENTENÇA: SIM

```

[[0, 'Operações (OP)', 'entity'], [1, 'ÁGATA', 'entity'], [2, 'Defesa', 'entity'], [3, 'segurança',
'entity'], [4, 'combate', 'entity'], [5, 'presença', 'entity'], [6, 'Estado', 'entity'], [7, 'fronteira',
'entity'], [8, 'integração', 'entity'], [9, 'Forças Armadas', 'entity'], [10, 'segurança pública',
'entity'], [11, 'bloqueio', 'entity'], [12, 'controle', 'entity'], [13, 'Ágata', 'entity'], [16,

```

```
'APROVEITAMENTO DO ÊXITO', 'entity'], [17, 'operação ofensiva', 'entity'], [18, 'ataque',
'entity'], [19, 'força', 'entity'], [20, 'movimento', 'entity']]
<built-in method count of list object at 0x7de389f286c0>
[[0, 'APROVEITAMENTO DO ÊXITO', 'associated_with', 'ataque'], [1, 'APROVEITAMENTO
DO ÊXITO', 'associated_with', 'força'], [2, 'APROVEITAMENTO DO ÊXITO',
'associated_with', 'movimento'], [3, 'operação ofensiva', 'type_of', 'APROVEITAMENTO DO
ÊXITO']]
```

TAREFA: NER

CORPUS:

PIPELINE: pt_core_news_lg

BATCH_SIZE = 128 (Linha: 14)

MAX_BATCH_ITEMS = 4096 (Linha: 48)

NAME = "neuralmind/bert-base-portuguese-cased" (Linha: 53)

DROPOUT = 0.2 (Linha: 92)

MAX_STEPS = 20000 (Linha: 95)

===== Training stats

=====

Language: pt

Training pipeline: transformer, ner

693 training docs

1618 evaluation docs

✓ No overlap between training and evaluation data

⚠ Low number of examples to train a new pipeline (693)

===== Vocab & Vectors

=====

i 31275 total word(s) in the data (5496 unique)

i No word vectors present in the package

===== Named Entity Recognition

=====

i 1 label(s)

0 missing value(s) (tokens with '-' label)

✓ Good amount of examples for all labels

✓ Examples without occurrences available for all labels

✓ No entities consisting of or starting/ending with whitespace

✓ No entities crossing sentence boundaries

===== Summary

=====

✓ 7 checks passed

⚠ 1 warning

===== Training pipeline

=====

i Pipeline: ['transformer', 'ner']

i Initial learn rate: 0.0

E # LOSS TRANS... LOSS NER ENTS_F ENTS_P ENTS_R SCORE

```
---
0 0 728.22 563.01 3.52 2.44 6.29 0.04
9 200 48014.91 57924.82 74.99 83.47 68.07 0.75
18 400 2974.91 5582.44 80.15 82.76 77.69 0.80
27 600 579.76 1996.47 79.65 80.15 79.14 0.80
37 800 338.30 1576.46 80.65 84.77 76.91 0.81
46 1000 266.20 1398.09 81.35 81.42 81.27 0.81
55 1200 184.86 1360.86 81.11 83.75 78.63 0.81
64 1400 95.47 1107.53 82.11 84.10 80.21 0.82
73 1600 36.20 1018.13 81.91 81.37 82.46 0.82
83 1800 87.46 1011.58 81.45 82.05 80.86 0.81
92 2000 74.86 918.66 81.41 81.41 81.42 0.81
101 2200 32.42 824.25 80.54 82.71 78.49 0.81
110 2400 74.83 795.33 82.57 83.01 82.13 0.83
120 2600 126.97 766.11 82.28 83.59 81.01 0.82
129 2800 50.65 648.56 82.10 81.33 82.88 0.82
138 3000 126.52 626.09 82.46 86.48 78.79 0.82
147 3200 53.15 517.93 82.84 84.60 81.14 0.83
157 3400 110.92 480.50 82.09 84.52 79.79 0.82
166 3600 109.15 402.07 83.02 84.39 81.69 0.83
175 3800 50.32 320.35 82.47 82.66 82.28 0.82
184 4000 149.87 339.93 82.76 85.25 80.42 0.83
194 4200 111.75 270.64 83.11 84.61 81.65 0.83
203 4400 47.31 203.94 83.02 83.59 82.46 0.83
212 4600 59.51 172.49 82.58 85.41 79.94 0.83
221 4800 111.44 147.17 83.47 86.85 80.33 0.83
230 5000 111.76 132.55 82.84 85.76 80.12 0.83
240 5200 231.64 127.54 82.61 84.78 80.55 0.83
249 5400 1397.40 115.56 83.48 86.25 80.89 0.83
258 5600 108.88 75.05 82.94 87.58 78.76 0.83
268 5800 28.02 45.21 83.21 86.35 80.29 0.83
277 6000 38.56 41.01 82.42 82.59 82.25 0.82
286 6200 63.05 50.37 83.05 86.24 80.08 0.83
295 6400 25.92 23.64 83.46 85.83 81.23 0.83
305 6600 146.92 32.93 83.10 84.32 81.92 0.83
314 6800 50.54 39.08 83.26 85.00 81.58 0.83
323 7000 25.11 20.52 82.19 85.34 79.27 0.82
```

✓ Saved pipeline to output directory

/content/drive/MyDrive/Doutorado/IDEA/NER/outputs/model-last

===== Results

=====

TOK 100.00

NER P 85.44

NER R 80.24
NER F 82.76
SPEED 5150

===== NER (per type)
=====

	P	R	F
entity	85.44	80.24	82.76

TAREFA: Relation Extraction

batch_size = 2000 (Linha: 19)
name = "neuralmind/bert-base-portuguese-cased" (Linha: 30)
threshold = 0.3 (Linha: 40)
max_length = 250 (Linha: 60)
max_steps = 1000 (Linha: 88)
size = 2000 (Linha: 99)
buffer = 256 (Linha: 100)

===== Training pipeline
=====

i Pipeline: ['transformer', 'relation_extractor']
i Initial learn rate: 0.0

E	#	LOSS TRANS...	LOSS RELAT...	REL_MICRO_P	REL_MICRO_R	REL_MICRO_F	SCORE
---	---	---------------	---------------	-------------	-------------	-------------	-------

0	0	6.13	2.86	1.21	98.73	2.38	0.02
4	100	174.11	71.18	89.60	89.32	89.46	0.89
9	200	0.08	2.46	90.64	89.06	89.84	0.90
13	300	0.10	2.11	94.99	87.89	91.31	0.91
18	400	0.10	1.79	90.54	89.70	90.12	0.90
23	500	0.06	1.12	92.05	90.61	91.33	0.91
27	600	0.03	0.92	93.20	90.06	91.60	0.92
32	700	0.01	0.77	92.50	90.89	91.69	0.92
36	800	0.01	0.65	92.40	90.69	91.54	0.92
41	900	0.01	0.65	93.07	90.72	91.88	0.92
46	1000	0.01	0.64	92.44	90.97	91.70	0.92

✓ Saved pipeline to output directory
training/model-last

Random baseline:

threshold 0.00	{'rel_micro_p': '0.98', 'rel_micro_r': '100.00', 'rel_micro_f': '1.94'}
threshold 0.05	{'rel_micro_p': '0.98', 'rel_micro_r': '94.62', 'rel_micro_f': '1.93'}
threshold 0.10	{'rel_micro_p': '0.97', 'rel_micro_r': '89.10', 'rel_micro_f': '1.92'}
threshold 0.20	{'rel_micro_p': '0.97', 'rel_micro_r': '78.99', 'rel_micro_f': '1.91'}
threshold 0.30	{'rel_micro_p': '0.98', 'rel_micro_r': '69.79', 'rel_micro_f': '1.93'}

threshold 0.40	{'rel_micro_p': '0.98', 'rel_micro_r': '60.17', 'rel_micro_f': '1.93'}
threshold 0.50	{'rel_micro_p': '0.99', 'rel_micro_r': '50.36', 'rel_micro_f': '1.93'}
threshold 0.60	{'rel_micro_p': '0.98', 'rel_micro_r': '39.81', 'rel_micro_f': '1.91'}
threshold 0.70	{'rel_micro_p': '0.97', 'rel_micro_r': '29.77', 'rel_micro_f': '1.89'}
threshold 0.80	{'rel_micro_p': '0.99', 'rel_micro_r': '20.31', 'rel_micro_f': '1.90'}
threshold 0.90	{'rel_micro_p': '0.99', 'rel_micro_r': '10.09', 'rel_micro_f': '1.80'}
threshold 0.99	{'rel_micro_p': '1.17', 'rel_micro_r': '1.19', 'rel_micro_f': '1.18'}
threshold 1.00	{'rel_micro_p': '0.67', 'rel_micro_r': '0.07', 'rel_micro_f': '0.13'}

Results of the trained model:

threshold 0.00	{'rel_micro_p': '0.98', 'rel_micro_r': '100.00', 'rel_micro_f': '1.94'}
threshold 0.05	{'rel_micro_p': '82.80', 'rel_micro_r': '89.12', 'rel_micro_f': '85.84'}
threshold 0.10	{'rel_micro_p': '87.02', 'rel_micro_r': '88.73', 'rel_micro_f': '87.87'}
threshold 0.20	{'rel_micro_p': '89.54', 'rel_micro_r': '88.49', 'rel_micro_f': '89.01'}
threshold 0.30	{'rel_micro_p': '90.65', 'rel_micro_r': '88.07', 'rel_micro_f': '89.34'}
threshold 0.40	{'rel_micro_p': '91.44', 'rel_micro_r': '87.61', 'rel_micro_f': '89.48'}
threshold 0.50	{'rel_micro_p': '91.97', 'rel_micro_r': '87.24', 'rel_micro_f': '89.54'}
threshold 0.60	{'rel_micro_p': '92.71', 'rel_micro_r': '86.84', 'rel_micro_f': '89.68'}
threshold 0.70	{'rel_micro_p': '93.42', 'rel_micro_r': '86.33', 'rel_micro_f': '89.73'}
threshold 0.80	{'rel_micro_p': '94.00', 'rel_micro_r': '85.42', 'rel_micro_f': '89.50'}
threshold 0.90	{'rel_micro_p': '94.86', 'rel_micro_r': '83.76', 'rel_micro_f': '88.97'}
threshold 0.99	{'rel_micro_p': '96.98', 'rel_micro_r': '77.85', 'rel_micro_f': '86.37'}
threshold 1.00	{'rel_micro_p': '98.46', 'rel_micro_r': '58.14', 'rel_micro_f': '73.11'}

CT_TEXTO = “2.2.5 A função de combate Inteligência é muito mais que a simples obtenção de dados e informações. É um processo contínuo que integra a análise da informação com o desenvolvimento das operações, de maneira que se possa visualizar e entender a situação. 2.2.6 Esta função de combate não inclui apenas o pessoal e os meios que a integram de forma específica. Dela também fazem parte todos aqueles que realizam, em determinado momento, de uma forma ou de outra, atividades próprias a ela. Todo militar é, assim, um meio de obtenção de dados em potencial (ESS – conceito do inglês “Every Soldier is a Sensor”). 2.2.7 A missão da função de combate Inteligência é apoiar o planejamento, a preparação, a execução e a avaliação de todos os tipos de operações. Portanto, o papel mais importante que desempenha é o de servir de base para o desenvolvimento das operações, apoiando o processo decisório, numa atividade contínua e dinâmica. 2.2.8 A função de combate inteligência necessita de uma configuração capaz de proporcionar estruturas específicas de inteligência e de comunicações a todos os níveis de planejamento. 2.2.9 A estrutura de inteligência deve incluir sistemas, procedimentos e organizações de inteligência capazes de gerar conhecimento de maneira oportuna. Uma estrutura de Tecnologia da Informação e das Comunicações (TIC) adequada complementa a função de combate. 2.2.10 A função de combate Inteligência, orientada pela definição das necessidades de inteligência, obtém os dados necessários por meio de um esforço de obtenção (adaptado ao escalão de emprego) e os analisa e integra, apoiando a manobra. 2.2.11 Uma vez orientada, elabora seu esforço de obtenção de dados, que consiste na

conjugação, no tempo e no espaço, dos sensores especializados e não especializados. A função de combate inteligência elabora e difunde respostas às necessidades de inteligência levantadas durante o exame de situação. O conhecimento elaborado é difundido, em seguida, por meio dos escalões subordinados e superiores, para todos aqueles em operações.”

```
LST_NER = [[0, 'FUNÇÃO DE COMBATE INTELIGÊNCIA', 'entity'], [1, 'combate', 'entity'], [2, 'operacional', 'entity'], [3, 'Comandante', 'entity'], [4, 'inteligência', 'entity'], [5, 'Inteligência', 'entity'], [6, 'Reconhecimento', 'entity'], [7, 'Vigilância', 'entity'], [8, 'Aquisição', 'entity'], [9, 'objetivo', 'entity'], [10, 'conhecimento', 'entity'], [11, 'Comando', 'entity'], [12, 'planejamento', 'entity'], [13, 'paz', 'entity'], [14, 'Ciclo de Inteligência', 'entity'], [15, 'Orientação', 'entity'], [16, 'Obtenção', 'entity'], [18, 'Atividade de Inteligência Militar Terrestre', 'entity'], [19, 'vigilância', 'entity'], [20, 'reconhecimento', 'entity'], [21, 'aquisição', 'entity'], [22, 'combate Inteligência', 'entity'], [23, 'obtenção', 'entity'], [24, 'processo', 'entity'], [25, 'análise da informação', 'entity'], [26, 'desenvolvimento', 'entity'], [28, 'meios', 'entity'], [30, 'missão', 'entity'], [34, 'preparação', 'entity'], [35, 'avaliação', 'entity'], [37, 'processo decisório', 'entity'], [38, 'atividade', 'entity'], [41, 'estrutura de inteligência', 'entity'], [44, 'Tecnologia da Informação', 'entity'], [48, 'esforço', 'entity'], [50, 'escalão', 'entity'], [51, 'emprego', 'entity']]
```

```
LST_RE = [[0, 'FUNÇÃO DE COMBATE INTELIGÊNCIA', 'associated_with', 'combate']]
```