

Judge System

Judge System, attacker ve defender'ın yazdığı cümlelerin doğruluğunu ve akıcılığını kontrol eder.

Adequacy check için, BERTurk Cased, post-response datasetinde eşleşen ve eşleşmeyen sample'ları sınıflandırmak için finetune edilmiştir.

Batch Size = 128, Learning Rate = 0.01

Epoch	Accuracy1	Accuracy2	Accuracy3	Ortalama	Std
3	51.5	50	48.75	50.08	1.12
5	49.75	51.75	51	50.83	0.82
10	52.25	51.25	49.75	51.08	1.03
20	50.25	53.5	46	49.92	3.07

Epoch = 10, Learning Rate = 0.01

Batch	Accuracy1	Accuracy2	Accuracy3	Ortalama	Std
32	49.25	50.25	51.25	50.25	0.82
64	52.25	51	52.25	51.83	0.59
128	52.25	51.25	49.75	51.08	1.03
256	55.25	54.25	51.5	53.67	1.59

Epoch = 10, Batch Size = 256

Learning Rate	Accuracy1	Accuracy2	Accuracy3	Ortalama	Std
0.1	49.25	-	-	-	-
0.01	55.25	54.25	51.5	53.67	1.59
0.001	53.25	50	53	52.08	1.48
0.0001	50	52	52.25	51.42	1.01

Learning rate 0.1'ken loss azalmıyor.

Genel olarak loss batch'lerde azalsa da accuracy %50'ler civarında kalıyor. Model, post'lar ve response'lar arasında bir ilişki olduğunu elimizdeki datasetinden (post_response_75.csv) öğrenemiyor.

Fluency check için Donanım Haber verisi GPT2-Turk modeline verilerek text üretme taskı için finetune edilmiştir.

Batch Size = 2, Learning Rate = 0.01

Epoch	Perplexity1	Perplexity2	Perplexity3	Ortalama	Std
1	20.06	19.89	20.37	20.11	0.2
3	21.01	20.7	21.09	20.93	0.17
5	21	20.63	20.66	20.76	0.17
10	23.79	21.32	21.33	22.15	1.16

Epoch = 10, Learning Rate = 0.01

Batch	Perplexity1	Perplexity2	Perplexity3	Ortalama	Std
2	23.79	21.32	21.33	22.15	1.16
4	22.45	21.54	21.31	21.77	0.49

Epoch = 10, Batch Size = 2

Learning Rate	Perplexity1	Perplexity2	Perplexity3	Ortalama	Std
0.1	54.95	357.64	55.77	156.12	142.5
0.01	23.79	21.32	21.33	22.15	1.16
0.005	20.08	19.36	20.56	20	0.49
0.001	21.09	21.16	21.7	21.32	0.27

Genel olarak GPT2 iyi çalışıyor. Paper’da perplexity değeri 35’in altında olan örneklerin akıcı olduğu kabul edildiği düşünüldüğünde test setlerin perplexity değerleri yeterince iyi.