HPO - Retriever Only - v1

Training a retriever model using the golf document as a target

Valentin Liévin

Setup

Searching for the best hyperparameters using N=636 trials with HyperOpt search alg. and Async HyperBand scheduler. Each model is trained using 4 GPUs and a training batch size of 24. The retriever model is implemented and trained as in Dense Passage Retrieval using the gold document as a target.

Results

BioBERT provides the best results, although the uncased SciBERT provides better top-1 test accuracy in average. Nonetheless, BioBERT outperforms SciBERT when considering the top-k test accuracy with $k > 1$. Dropout=0 and lr=0.0001 are also important factors.

The best model scores 11% top-1 test accuracy and 36.3% top-10 test accuracy using the following hyperparameters:

- hidden_size=128
- bert=BioBERT
- lr =1e-4
- weight_decay=1e-3
- dropout=0

The best models are reported bellow:

Discussion

Results should be compared with the BM25 baseline.

Data

Top parameters

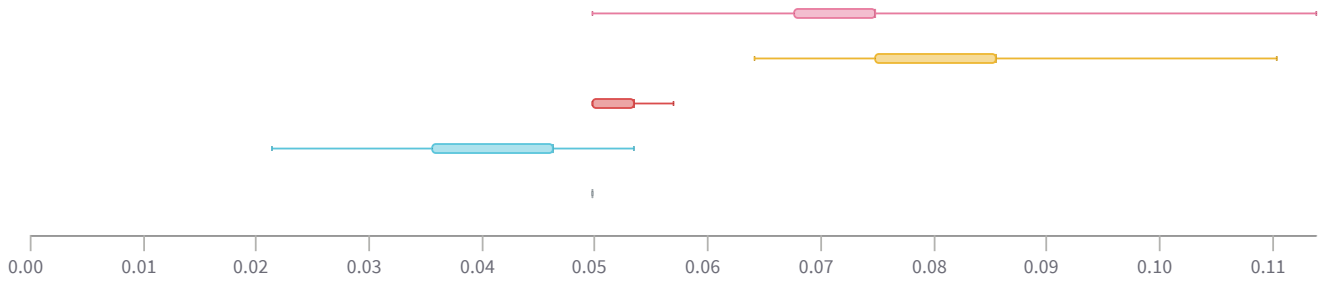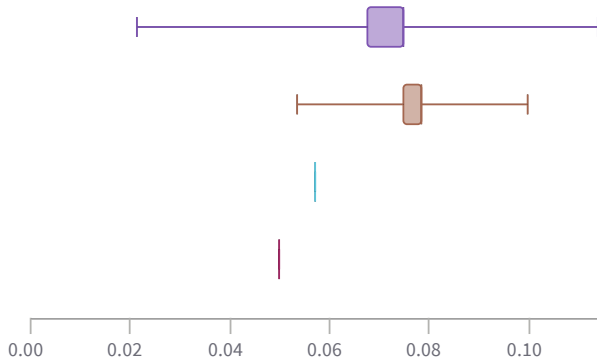| Name (692 visualized)	Runtime	model/hidd	model/weig	model/lr	model/drop	test/top100
summer-thunder-1112	31m 32s	128	0.00001	0.00001	0	0.8043
comfy-yogurt-459	22m 23s	128	0	0.00001	0	0.7758
balmy-forest-684	22m 41s	128	0.00001	0.00001	0	0.7794
restful-microwave-1125	20m 48s	128	0.001	0.00003	0	0.7794
zesty-snowflake-744	32m 56s	128	0.00001	0.00001	0	0.7865
revived-wave-251	21m 57s	128	0.00001	0.00001	0	0.7758
curious-leaf-532	22m 52s	128	0.00001	0.00001	0	0.7936
vivid-sponge-582	21m 46s	128	0.00001	0.00001	0	0.7865
different-lion-1124	20m 42s	128	0	0.00003	0	0.8078
grateful-plasma-910	19m 23s	128	0.001	0.00003	0	0.7794

Run set 692
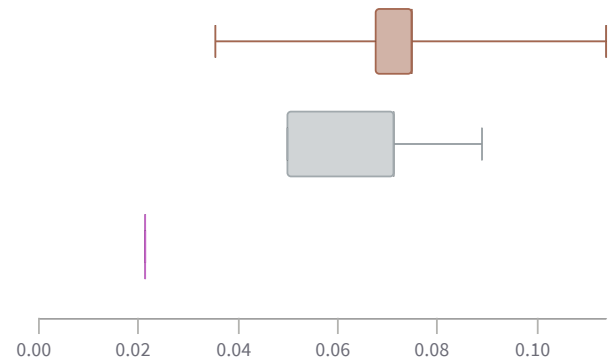
1-10 ▾ of 692

Top-1 accuracy
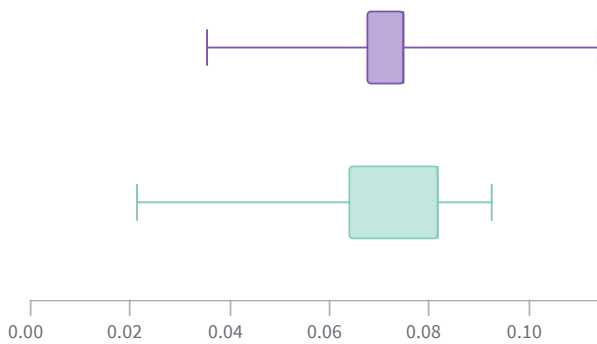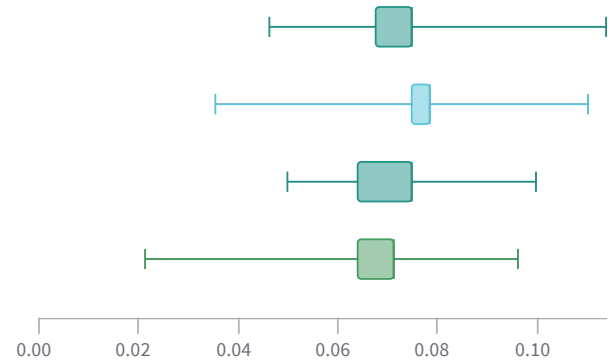
Accuracy vs. BERT type

Accuracy vs. learning rate
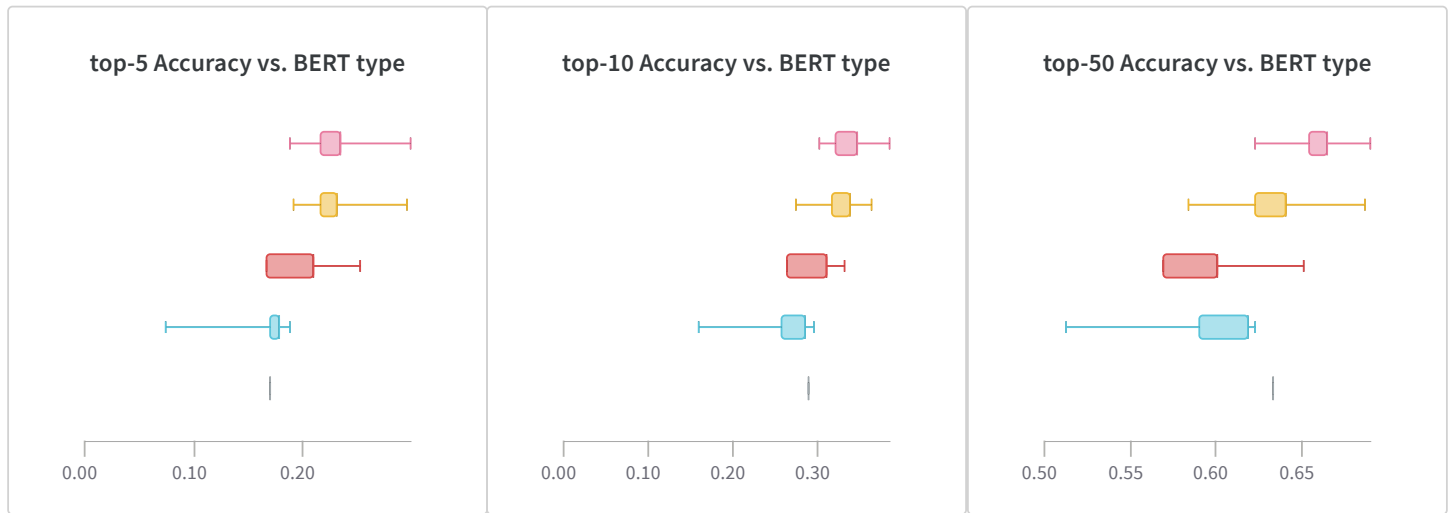
Accuracy vs. dropout

Accuracy vs. hidden_size

Accuracy vs. weight decay

▾ top-k accuracy vs. BERT

Created with ❤️ on Weights & Biases.

https://wandb.ai/vlievin/findzebra-qa/reports/HPO-Retriever-Only-v1--Vmlldzo3NzgyMjc