

Document QA model

Environment:

Tensorflow==1.3.0

Cuda==8.0

Cudnn==6.0

```
(python35) F:\bi-att-flow-master>python squad/evaluate-v1.1.py C:\Users\Administrator\data
{"exact_match": 66.59413434247871, "f1": 76.31287252963799}
```

```
TrainParams(SerializableOptimizer("Adadelta", dict(learning_rate=1.0)),
            ema=0.999, max_checkpoints_to_keep=3, async_encoding=10,
            num_epochs=n_epochs, log_period=30, eval_period=1200, save_period=1200,
            eval_samples=dict(dev=None, train=8000))
```

Paragraph level:

Restoring variables

Restoring EMA variables

Begin evaluation

[illegible]

Metric

b17/accuracy 0.6854304635761589

| | |
|--------|-------------------|
| b17/f1 | 0.789331924593681 |
|--------|-------------------|

b17/text-em 0.7151371807000946

b17/text-f1 0.8047825020162882

```
ubuntu@ip-172-31-79-175:~/document-qa$ python3.6 docqa/squad/squad_official_evaluation.py ~/data/squad/dev-v1.1.json output_para.json
{"exact_match": 71.51371807000946, "f1": 80.47825020162885}
```

Saving result

| N Paragraphs | EM | F1 |
|--------------|------|------|
| 10 | 0.00 | 0.00 |
| 20 | 0.00 | 0.00 |
| 30 | 0.00 | 0.00 |
| 40 | 0.00 | 0.00 |
| 50 | 0.00 | 0.00 |
| 60 | 0.00 | 0.00 |
| 70 | 0.00 | 0.00 |
| 80 | 0.00 | 0.00 |
| 90 | 0.00 | 0.00 |
| 100 | 0.00 | 0.00 |

| | | |
|---|--------|--------|
| 1 | 0.5608 | 0.6380 |
|---|--------|--------|

| | | |
|---|--------|--------|
| 2 | 0.5932 | 0.6717 |
|---|--------|--------|

| | | |
|---|--------|--------|
| 3 | 0.5877 | 0.6648 |
|---|--------|--------|

| | | |
|---|--------|--------|
| 4 | 0.5753 | 0.6508 |
|---|--------|--------|

| | | |
|---|--------|--------|
| 5 | 0.5667 | 0.6402 |
|---|--------|--------|

| | | |
|---|--------|--------|
| 5 | 0.5507 | 0.5402 |
| 6 | 0.5607 | 0.6332 |

| | | |
|---|--------|--------|
| 6 | 0.5557 | 0.5557 |
| 7 | 0.5557 | 0.6274 |

| | | |
|---|--------|--------|
| 7 | 0.5557 | 0.6274 |
| 8 | 0.5500 | 0.6214 |

| | | |
|---|--------|--------|
| 9 | 0.5333 | 0.6214 |
| 9 | 0.5423 | 0.6128 |

| | | |
|----|--------|--------|
| 9 | 0.5129 | 0.5129 |
| 10 | 0.5379 | 0.6083 |

| | |
|----|---------------|
| 10 | 0.5379 0.6033 |
| 11 | 0.5344 0.6048 |

| | | |
|----|--------|--------|
| 11 | 0.5544 | 0.5543 |
| 12 | 0.5292 | 0.5993 |

| | |
|----|---------------|
| 12 | 0.5252 0.5555 |
| 13 | 0.5271 0.5961 |

| | |
|----|---------------|
| 13 | 0.5271 0.5901 |
| 14 | 0.5236 0.5925 |

| | |
|----|---------------|
| 14 | 0.5230 0.5929 |
| 15 | 0.5215 0.5897 |

| | | |
|----|--------|--------|
| 15 | 0.5215 | 0.5897 |
| 16 | 0.5185 | 0.5863 |

| | | |
|----|--------|--------|
| 17 | 0.5158 | 0.5832 |
| 18 | 0.5132 | 0.5803 |
| 19 | 0.5114 | 0.5785 |
| 20 | 0.5096 | 0.5766 |
| 21 | 0.5073 | 0.5740 |
| 22 | 0.5048 | 0.5710 |
| 23 | 0.5031 | 0.5691 |
| 24 | 0.5013 | 0.5672 |
| 25 | 0.4992 | 0.5651 |
| 26 | 0.4981 | 0.5639 |
| 27 | 0.4969 | 0.5625 |
| 28 | 0.4953 | 0.5607 |
| 29 | 0.4939 | 0.5594 |
| 30 | 0.4930 | 0.5584 |
| 31 | 0.4916 | 0.5569 |
| 32 | 0.4903 | 0.5555 |
| 33 | 0.4886 | 0.5538 |
| 34 | 0.4871 | 0.5519 |
| 35 | 0.4870 | 0.5517 |
| 36 | 0.4861 | 0.5506 |
| 37 | 0.4847 | 0.5492 |
| 38 | 0.4834 | 0.5478 |
| 39 | 0.4818 | 0.5459 |
| 40 | 0.4811 | 0.5450 |
| 41 | 0.4804 | 0.5442 |
| 42 | 0.4800 | 0.5437 |
| 43 | 0.4794 | 0.5431 |
| 44 | 0.4791 | 0.5425 |
| 45 | 0.4786 | 0.5419 |
| 46 | 0.4781 | 0.5415 |
| 47 | 0.4779 | 0.5412 |
| 48 | 0.4774 | 0.5406 |
| 49 | 0.4767 | 0.5398 |
| 50 | 0.4767 | 0.5396 |
| 51 | 0.4762 | 0.5390 |
| 52 | 0.4759 | 0.5388 |
| 53 | 0.4754 | 0.5383 |
| 54 | 0.4747 | 0.5377 |
| 55 | 0.4744 | 0.5374 |
| 56 | 0.4740 | 0.5369 |
| 57 | 0.4738 | 0.5368 |
| 58 | 0.4734 | 0.5364 |
| 59 | 0.4732 | 0.5361 |
| 60 | 0.4731 | 0.5360 |
| 61 | 0.4731 | 0.5359 |
| 62 | 0.4731 | 0.5360 |
| 63 | 0.4729 | 0.5357 |
| 64 | 0.4728 | 0.5355 |
| 65 | 0.4727 | 0.5355 |
| 66 | 0.4727 | 0.5354 |
| 67 | 0.4727 | 0.5354 |
| 68 | 0.4725 | 0.5352 |
| 69 | 0.4722 | 0.5349 |
| 70 | 0.4721 | 0.5348 |
| 71 | 0.4721 | 0.5348 |
| 72 | 0.4718 | 0.5345 |
| 73 | 0.4718 | 0.5344 |

| | | |
|----|--------|--------|
| 74 | 0.4716 | 0.5342 |
| 75 | 0.4717 | 0.5343 |
| 76 | 0.4714 | 0.5339 |
| 77 | 0.4714 | 0.5339 |
| 78 | 0.4713 | 0.5337 |
| 79 | 0.4714 | 0.5338 |
| 80 | 0.4711 | 0.5336 |
| 81 | 0.4712 | 0.5337 |
| 82 | 0.4712 | 0.5336 |
| 83 | 0.4711 | 0.5333 |
| 84 | 0.4710 | 0.5331 |
| 85 | 0.4709 | 0.5330 |
| 86 | 0.4706 | 0.5327 |
| 87 | 0.4705 | 0.5326 |
| 88 | 0.4705 | 0.5326 |
| 89 | 0.4706 | 0.5328 |
| 90 | 0.4705 | 0.5327 |
| 91 | 0.4704 | 0.5326 |
| 92 | 0.4704 | 0.5326 |
| 93 | 0.4704 | 0.5326 |
| 94 | 0.4704 | 0.5326 |
| 95 | 0.4704 | 0.5326 |
| 96 | 0.4704 | 0.5326 |
| 97 | 0.4704 | 0.5325 |
| 98 | 0.4704 | 0.5325 |

ubuntu@ip-172-31-79-175:~/document-qa\$ python3.6 docqa/eval/ranked_scores.py output_para.csv

Loading answers..

Computing ranks...

Rank answers_0_em answers_0_f1

| | | |
|----|--------|--------|
| 1 | 0.5608 | 0.6380 |
| 2 | 0.5932 | 0.6717 |
| 3 | 0.5877 | 0.6648 |
| 4 | 0.5753 | 0.6508 |
| 5 | 0.5667 | 0.6402 |
| 6 | 0.5607 | 0.6332 |
| 7 | 0.5557 | 0.6274 |
| 8 | 0.5500 | 0.6214 |
| 9 | 0.5423 | 0.6128 |
| 10 | 0.5379 | 0.6083 |
| 11 | 0.5344 | 0.6048 |
| 12 | 0.5292 | 0.5993 |
| 13 | 0.5271 | 0.5961 |
| 14 | 0.5236 | 0.5925 |
| 15 | 0.5215 | 0.5897 |
| 16 | 0.5185 | 0.5863 |
| 17 | 0.5158 | 0.5832 |
| 18 | 0.5132 | 0.5803 |
| 19 | 0.5114 | 0.5785 |
| 20 | 0.5096 | 0.5766 |
| 21 | 0.5073 | 0.5740 |
| 22 | 0.5048 | 0.5710 |
| 23 | 0.5031 | 0.5691 |
| 24 | 0.5013 | 0.5672 |
| 25 | 0.4992 | 0.5651 |
| 26 | 0.4981 | 0.5639 |
| 27 | 0.4969 | 0.5625 |

| | | |
|----|--------|--------|
| 28 | 0.4953 | 0.5607 |
| 29 | 0.4939 | 0.5594 |
| 30 | 0.4930 | 0.5584 |
| 31 | 0.4916 | 0.5569 |
| 32 | 0.4903 | 0.5555 |
| 33 | 0.4886 | 0.5538 |
| 34 | 0.4871 | 0.5519 |
| 35 | 0.4870 | 0.5517 |
| 36 | 0.4861 | 0.5506 |
| 37 | 0.4847 | 0.5492 |
| 38 | 0.4834 | 0.5478 |
| 39 | 0.4818 | 0.5459 |
| 40 | 0.4811 | 0.5450 |
| 41 | 0.4804 | 0.5442 |
| 42 | 0.4800 | 0.5437 |
| 43 | 0.4794 | 0.5431 |
| 44 | 0.4791 | 0.5425 |
| 45 | 0.4786 | 0.5419 |
| 46 | 0.4781 | 0.5415 |
| 47 | 0.4779 | 0.5412 |
| 48 | 0.4774 | 0.5406 |
| 49 | 0.4767 | 0.5398 |
| 50 | 0.4767 | 0.5396 |
| 51 | 0.4762 | 0.5390 |
| 52 | 0.4759 | 0.5388 |
| 53 | 0.4754 | 0.5383 |
| 54 | 0.4747 | 0.5377 |
| 55 | 0.4744 | 0.5374 |
| 56 | 0.4740 | 0.5369 |
| 57 | 0.4738 | 0.5368 |
| 58 | 0.4734 | 0.5364 |
| 59 | 0.4732 | 0.5361 |
| 60 | 0.4731 | 0.5360 |
| 61 | 0.4731 | 0.5359 |
| 62 | 0.4731 | 0.5360 |
| 63 | 0.4729 | 0.5357 |
| 64 | 0.4728 | 0.5355 |
| 65 | 0.4727 | 0.5355 |
| 66 | 0.4727 | 0.5354 |
| 67 | 0.4727 | 0.5354 |
| 68 | 0.4725 | 0.5352 |
| 69 | 0.4722 | 0.5349 |
| 70 | 0.4721 | 0.5348 |
| 71 | 0.4721 | 0.5348 |
| 72 | 0.4718 | 0.5345 |
| 73 | 0.4718 | 0.5344 |
| 74 | 0.4716 | 0.5342 |
| 75 | 0.4717 | 0.5343 |
| 76 | 0.4714 | 0.5339 |
| 77 | 0.4714 | 0.5339 |
| 78 | 0.4713 | 0.5337 |
| 79 | 0.4714 | 0.5338 |
| 80 | 0.4711 | 0.5336 |
| 81 | 0.4712 | 0.5337 |
| 82 | 0.4712 | 0.5336 |
| 83 | 0.4711 | 0.5333 |
| 84 | 0.4710 | 0.5331 |

| | | |
|----|--------|--------|
| 85 | 0.4709 | 0.5330 |
| 86 | 0.4706 | 0.5327 |
| 87 | 0.4705 | 0.5326 |
| 88 | 0.4705 | 0.5326 |
| 89 | 0.4706 | 0.5328 |
| 90 | 0.4705 | 0.5327 |
| 91 | 0.4704 | 0.5326 |
| 92 | 0.4704 | 0.5326 |
| 93 | 0.4704 | 0.5326 |
| 94 | 0.4704 | 0.5326 |
| 95 | 0.4704 | 0.5326 |
| 96 | 0.4704 | 0.5326 |
| 97 | 0.4704 | 0.5325 |
| 98 | 0.4704 | 0.5325 |

Shared_norm method

```
python3.6 docqa/eval/squad_eval.py -o output.json -c dev shared-norm-output-0412-145610
```

```
Restoring variables
Restoring EMA variables
Begin evaluation
dev: 100%|██████████████████████████████████████| 53/53 [00:24<00:00, 2.15it/s]
Metric
b17/accuracy 0.6829706717123936
b17/f1        0.7913130171392136
b17/text-em   0.7143803216650899
b17/text-f1   0.8066652931379928
```

```
[ubuntu@ip-172-31-79-175:~/document-qa$ python3.6 docqa/squad/squad_official_evaluation.py ~/data/squad/dev-v1.1.json output.json]
{"exact_match": 71.43803216650899, "f1": 80.66652931379933}
```

| Saving result | | |
|---------------|--------|--------|
| N Paragraphs | EM | F1 |
| 1 | 0.5597 | 0.6387 |
| 2 | 0.6198 | 0.7017 |
| 3 | 0.6315 | 0.7137 |
| 4 | 0.6371 | 0.7196 |
| 5 | 0.6410 | 0.7236 |
| 6 | 0.6431 | 0.7257 |
| 7 | 0.6454 | 0.7279 |
| 8 | 0.6458 | 0.7283 |
| 9 | 0.6450 | 0.7275 |
| 10 | 0.6448 | 0.7275 |
| 11 | 0.6453 | 0.7278 |
| 12 | 0.6454 | 0.7282 |
| 13 | 0.6458 | 0.7284 |
| 14 | 0.6455 | 0.7283 |
| 15 | 0.6450 | 0.7279 |
| 16 | 0.6449 | 0.7277 |
| 17 | 0.6444 | 0.7270 |
| 18 | 0.6442 | 0.7266 |
| 19 | 0.6442 | 0.7268 |
| 20 | 0.6444 | 0.7268 |
| 21 | 0.6442 | 0.7266 |
| 22 | 0.6441 | 0.7264 |

| | |
|----|---------------|
| 23 | 0.6436 0.7259 |
| 24 | 0.6434 0.7257 |
| 25 | 0.6433 0.7256 |
| 26 | 0.6434 0.7258 |
| 27 | 0.6436 0.7260 |
| 28 | 0.6435 0.7260 |
| 29 | 0.6430 0.7255 |
| 30 | 0.6431 0.7256 |
| 31 | 0.6432 0.7257 |
| 32 | 0.6432 0.7258 |
| 33 | 0.6433 0.7259 |
| 34 | 0.6431 0.7256 |
| 35 | 0.6430 0.7255 |
| 36 | 0.6430 0.7254 |
| 37 | 0.6430 0.7254 |
| 38 | 0.6429 0.7253 |
| 39 | 0.6430 0.7255 |
| 40 | 0.6430 0.7253 |
| 41 | 0.6430 0.7254 |
| 42 | 0.6430 0.7253 |
| 43 | 0.6430 0.7254 |
| 44 | 0.6430 0.7254 |
| 45 | 0.6431 0.7256 |
| 46 | 0.6432 0.7257 |
| 47 | 0.6432 0.7257 |
| 48 | 0.6431 0.7256 |
| 49 | 0.6430 0.7254 |
| 50 | 0.6430 0.7253 |
| 51 | 0.6430 0.7253 |
| 52 | 0.6430 0.7253 |
| 53 | 0.6430 0.7252 |
| 54 | 0.6430 0.7253 |
| 55 | 0.6431 0.7254 |
| 56 | 0.6430 0.7253 |
| 57 | 0.6430 0.7253 |
| 58 | 0.6430 0.7252 |
| 59 | 0.6430 0.7251 |
| 60 | 0.6430 0.7251 |
| 61 | 0.6430 0.7251 |
| 62 | 0.6430 0.7251 |
| 63 | 0.6430 0.7252 |
| 64 | 0.6430 0.7251 |
| 65 | 0.6430 0.7252 |
| 66 | 0.6430 0.7251 |
| 67 | 0.6430 0.7252 |
| 68 | 0.6430 0.7252 |
| 69 | 0.6430 0.7251 |
| 70 | 0.6430 0.7251 |
| 71 | 0.6430 0.7251 |
| 72 | 0.6430 0.7251 |
| 73 | 0.6430 0.7251 |
| 74 | 0.6430 0.7251 |
| 75 | 0.6431 0.7253 |
| 76 | 0.6431 0.7253 |
| 77 | 0.6433 0.7254 |
| 78 | 0.6433 0.7254 |
| 79 | 0.6434 0.7255 |

| | | |
|----|--------|--------|
| 80 | 0.6432 | 0.7254 |
| 81 | 0.6433 | 0.7255 |
| 82 | 0.6434 | 0.7255 |
| 83 | 0.6433 | 0.7254 |
| 84 | 0.6433 | 0.7254 |
| 85 | 0.6433 | 0.7254 |
| 86 | 0.6433 | 0.7254 |
| 87 | 0.6433 | 0.7254 |
| 88 | 0.6432 | 0.7253 |
| 89 | 0.6432 | 0.7253 |
| 90 | 0.6432 | 0.7253 |
| 91 | 0.6431 | 0.7253 |
| 92 | 0.6432 | 0.7254 |
| 93 | 0.6432 | 0.7254 |
| 94 | 0.6432 | 0.7254 |
| 95 | 0.6432 | 0.7254 |
| 96 | 0.6432 | 0.7254 |
| 97 | 0.6432 | 0.7254 |
| 98 | 0.6432 | 0.7254 |

ubuntu@ip-172-31-79-175:~/document-qa\$ python3.6 docqa/eval/ranked_scores.py output_shared_norm.csv

Loading answers..

Computing ranks...

Rank answers_0_em answers_0_f1

| | | |
|----|--------|--------|
| 1 | 0.5597 | 0.6387 |
| 2 | 0.6198 | 0.7017 |
| 3 | 0.6315 | 0.7137 |
| 4 | 0.6371 | 0.7196 |
| 5 | 0.6410 | 0.7236 |
| 6 | 0.6431 | 0.7257 |
| 7 | 0.6454 | 0.7279 |
| 8 | 0.6458 | 0.7283 |
| 9 | 0.6450 | 0.7275 |
| 10 | 0.6448 | 0.7275 |
| 11 | 0.6453 | 0.7278 |
| 12 | 0.6454 | 0.7282 |
| 13 | 0.6458 | 0.7284 |
| 14 | 0.6455 | 0.7283 |
| 15 | 0.6450 | 0.7279 |
| 16 | 0.6449 | 0.7277 |
| 17 | 0.6444 | 0.7270 |
| 18 | 0.6442 | 0.7266 |
| 19 | 0.6442 | 0.7268 |
| 20 | 0.6444 | 0.7268 |
| 21 | 0.6442 | 0.7266 |
| 22 | 0.6441 | 0.7264 |
| 23 | 0.6436 | 0.7259 |
| 24 | 0.6434 | 0.7257 |
| 25 | 0.6433 | 0.7256 |
| 26 | 0.6434 | 0.7258 |
| 27 | 0.6436 | 0.7260 |
| 28 | 0.6435 | 0.7260 |
| 29 | 0.6430 | 0.7255 |
| 30 | 0.6431 | 0.7256 |
| 31 | 0.6432 | 0.7257 |
| 32 | 0.6432 | 0.7258 |
| 33 | 0.6433 | 0.7259 |
| 34 | 0.6431 | 0.7256 |

| | | |
|----|--------|--------|
| 35 | 0.6430 | 0.7255 |
| 36 | 0.6430 | 0.7254 |
| 37 | 0.6430 | 0.7254 |
| 38 | 0.6429 | 0.7253 |
| 39 | 0.6430 | 0.7255 |
| 40 | 0.6430 | 0.7253 |
| 41 | 0.6430 | 0.7254 |
| 42 | 0.6430 | 0.7253 |
| 43 | 0.6430 | 0.7254 |
| 44 | 0.6430 | 0.7254 |
| 45 | 0.6431 | 0.7256 |
| 46 | 0.6432 | 0.7257 |
| 47 | 0.6432 | 0.7257 |
| 48 | 0.6431 | 0.7256 |
| 49 | 0.6430 | 0.7254 |
| 50 | 0.6430 | 0.7253 |
| 51 | 0.6430 | 0.7253 |
| 52 | 0.6430 | 0.7253 |
| 53 | 0.6430 | 0.7252 |
| 54 | 0.6430 | 0.7253 |
| 55 | 0.6431 | 0.7254 |
| 56 | 0.6430 | 0.7253 |
| 57 | 0.6430 | 0.7253 |
| 58 | 0.6430 | 0.7252 |
| 59 | 0.6430 | 0.7251 |
| 60 | 0.6430 | 0.7251 |
| 61 | 0.6430 | 0.7251 |
| 62 | 0.6430 | 0.7251 |
| 63 | 0.6430 | 0.7252 |
| 64 | 0.6430 | 0.7251 |
| 65 | 0.6430 | 0.7252 |
| 66 | 0.6430 | 0.7251 |
| 67 | 0.6430 | 0.7252 |
| 68 | 0.6430 | 0.7252 |
| 69 | 0.6430 | 0.7251 |
| 70 | 0.6430 | 0.7251 |
| 71 | 0.6430 | 0.7251 |
| 72 | 0.6430 | 0.7251 |
| 73 | 0.6430 | 0.7251 |
| 74 | 0.6430 | 0.7251 |
| 75 | 0.6431 | 0.7253 |
| 76 | 0.6431 | 0.7253 |
| 77 | 0.6433 | 0.7254 |
| 78 | 0.6433 | 0.7254 |
| 79 | 0.6434 | 0.7255 |
| 80 | 0.6432 | 0.7254 |
| 81 | 0.6433 | 0.7255 |
| 82 | 0.6434 | 0.7255 |
| 83 | 0.6433 | 0.7254 |
| 84 | 0.6433 | 0.7254 |
| 85 | 0.6433 | 0.7254 |
| 86 | 0.6433 | 0.7254 |
| 87 | 0.6433 | 0.7254 |
| 88 | 0.6432 | 0.7253 |
| 89 | 0.6432 | 0.7253 |
| 90 | 0.6432 | 0.7253 |
| 91 | 0.6431 | 0.7253 |

| | | |
|----|--------|--------|
| 92 | 0.6432 | 0.7254 |
| 93 | 0.6432 | 0.7254 |
| 94 | 0.6432 | 0.7254 |
| 95 | 0.6432 | 0.7254 |
| 96 | 0.6432 | 0.7254 |
| 97 | 0.6432 | 0.7254 |
| 98 | 0.6432 | 0.7254 |

Merge

```
Restoring variables
Restoring EMA variables
Begin evaluation
dev: 100%|██████████████████████████████████████| 53/53 [00:26<00:00, 2.02it/s]
Metric
b17/accuracy 0.6754966887417219
b17/f1        0.7834783883660766
b17/text-em   0.7063386944181647
b17/text-f1   0.7988083019293953

ubuntu@ip-172-31-79-175:~/document-qa$ python3.6 docqa/squad/squad_official_eval_
uation.py ~/data/squad/dev-v1.1.json output.json
{"exact_match": 70.63386944181646, "f1": 79.8808301929396}
```

```
Restoring variables
Restoring EMA variables
Begin evaluation
dev: 100%|██████████| 2681/2681 [09:53<00:00, 4.52it/s]
scoring: 100%|██████████| 536115/536115 [01:13<00:00, 7298.93it/s]
Saving result
N Paragraphs EM   F1
1      0.5547 0.6348
2      0.6002 0.6833
3      0.6040 0.6858
4      0.6019 0.6829
5      0.5996 0.6802
6      0.5960 0.6757
7      0.5939 0.6735
8      0.5906 0.6695
9      0.5866 0.6648
10     0.5823 0.6609
11     0.5806 0.6585
12     0.5772 0.6549
13     0.5760 0.6531
14     0.5743 0.6513
15     0.5733 0.6504
16     0.5711 0.6481
17     0.5695 0.6460
18     0.5676 0.6439
19     0.5672 0.6432
20     0.5665 0.6423
21     0.5648 0.6403
22     0.5629 0.6383
23     0.5615 0.6368
24     0.5605 0.6358
25     0.5597 0.6346
26     0.5581 0.6330
27     0.5571 0.6316
```

| | | |
|----|--------|--------|
| 28 | 0.5567 | 0.6309 |
| 29 | 0.5553 | 0.6293 |
| 30 | 0.5536 | 0.6272 |
| 31 | 0.5526 | 0.6260 |
| 32 | 0.5518 | 0.6250 |
| 33 | 0.5509 | 0.6239 |
| 34 | 0.5503 | 0.6232 |
| 35 | 0.5495 | 0.6222 |
| 36 | 0.5488 | 0.6214 |
| 37 | 0.5474 | 0.6198 |
| 38 | 0.5468 | 0.6192 |
| 39 | 0.5468 | 0.6191 |
| 40 | 0.5465 | 0.6187 |
| 41 | 0.5462 | 0.6183 |
| 42 | 0.5459 | 0.6179 |
| 43 | 0.5455 | 0.6175 |
| 44 | 0.5450 | 0.6169 |
| 45 | 0.5448 | 0.6167 |
| 46 | 0.5449 | 0.6167 |
| 47 | 0.5447 | 0.6165 |
| 48 | 0.5443 | 0.6159 |
| 49 | 0.5439 | 0.6154 |
| 50 | 0.5433 | 0.6149 |
| 51 | 0.5433 | 0.6149 |
| 52 | 0.5431 | 0.6146 |
| 53 | 0.5430 | 0.6145 |
| 54 | 0.5427 | 0.6143 |
| 55 | 0.5427 | 0.6142 |
| 56 | 0.5426 | 0.6140 |
| 57 | 0.5425 | 0.6139 |
| 58 | 0.5420 | 0.6134 |
| 59 | 0.5418 | 0.6131 |
| 60 | 0.5415 | 0.6128 |
| 61 | 0.5414 | 0.6127 |
| 62 | 0.5412 | 0.6123 |
| 63 | 0.5408 | 0.6119 |
| 64 | 0.5404 | 0.6116 |
| 65 | 0.5399 | 0.6109 |
| 66 | 0.5397 | 0.6108 |
| 67 | 0.5396 | 0.6107 |
| 68 | 0.5395 | 0.6105 |
| 69 | 0.5395 | 0.6104 |
| 70 | 0.5395 | 0.6104 |
| 71 | 0.5395 | 0.6104 |
| 72 | 0.5395 | 0.6103 |
| 73 | 0.5395 | 0.6103 |
| 74 | 0.5393 | 0.6101 |
| 75 | 0.5391 | 0.6099 |
| 76 | 0.5390 | 0.6098 |
| 77 | 0.5390 | 0.6099 |
| 78 | 0.5388 | 0.6096 |
| 79 | 0.5389 | 0.6097 |
| 80 | 0.5386 | 0.6095 |
| 81 | 0.5387 | 0.6096 |
| 82 | 0.5385 | 0.6094 |
| 83 | 0.5383 | 0.6092 |
| 84 | 0.5382 | 0.6091 |

| | | |
|----|--------|--------|
| 85 | 0.5378 | 0.6087 |
| 86 | 0.5378 | 0.6087 |
| 87 | 0.5376 | 0.6084 |
| 88 | 0.5375 | 0.6084 |
| 89 | 0.5371 | 0.6080 |
| 90 | 0.5370 | 0.6079 |
| 91 | 0.5366 | 0.6075 |
| 92 | 0.5365 | 0.6074 |
| 93 | 0.5364 | 0.6073 |
| 94 | 0.5363 | 0.6072 |
| 95 | 0.5363 | 0.6072 |
| 96 | 0.5362 | 0.6072 |
| 97 | 0.5362 | 0.6072 |
| 98 | 0.5360 | 0.6070 |

python3.6 docqa/eval/ranked_scores.py output-merge.csv

Loading answers..

Computing ranks...

Rank answers_0_em answers_0_f1

| | | |
|----|--------|--------|
| 1 | 0.5547 | 0.6348 |
| 2 | 0.6002 | 0.6833 |
| 3 | 0.6040 | 0.6858 |
| 4 | 0.6019 | 0.6829 |
| 5 | 0.5996 | 0.6802 |
| 6 | 0.5960 | 0.6757 |
| 7 | 0.5939 | 0.6735 |
| 8 | 0.5906 | 0.6695 |
| 9 | 0.5866 | 0.6648 |
| 10 | 0.5823 | 0.6609 |
| 11 | 0.5806 | 0.6585 |
| 12 | 0.5772 | 0.6549 |
| 13 | 0.5760 | 0.6531 |
| 14 | 0.5743 | 0.6513 |
| 15 | 0.5733 | 0.6504 |
| 16 | 0.5711 | 0.6481 |
| 17 | 0.5695 | 0.6460 |
| 18 | 0.5676 | 0.6439 |
| 19 | 0.5672 | 0.6432 |
| 20 | 0.5665 | 0.6423 |
| 21 | 0.5648 | 0.6403 |
| 22 | 0.5629 | 0.6383 |
| 23 | 0.5615 | 0.6368 |
| 24 | 0.5605 | 0.6358 |
| 25 | 0.5597 | 0.6346 |
| 26 | 0.5581 | 0.6330 |
| 27 | 0.5571 | 0.6316 |
| 28 | 0.5567 | 0.6309 |
| 29 | 0.5553 | 0.6293 |
| 30 | 0.5536 | 0.6272 |
| 31 | 0.5526 | 0.6260 |
| 32 | 0.5518 | 0.6250 |
| 33 | 0.5509 | 0.6239 |
| 34 | 0.5503 | 0.6232 |
| 35 | 0.5495 | 0.6222 |
| 36 | 0.5488 | 0.6214 |
| 37 | 0.5474 | 0.6198 |
| 38 | 0.5468 | 0.6192 |
| 39 | 0.5468 | 0.6191 |

| | | |
|----|--------|--------|
| 40 | 0.5465 | 0.6187 |
| 41 | 0.5462 | 0.6183 |
| 42 | 0.5459 | 0.6179 |
| 43 | 0.5455 | 0.6175 |
| 44 | 0.5450 | 0.6169 |
| 45 | 0.5448 | 0.6167 |
| 46 | 0.5449 | 0.6167 |
| 47 | 0.5447 | 0.6165 |
| 48 | 0.5443 | 0.6159 |
| 49 | 0.5439 | 0.6154 |
| 50 | 0.5433 | 0.6149 |
| 51 | 0.5433 | 0.6149 |
| 52 | 0.5431 | 0.6146 |
| 53 | 0.5430 | 0.6145 |
| 54 | 0.5427 | 0.6143 |
| 55 | 0.5427 | 0.6142 |
| 56 | 0.5426 | 0.6140 |
| 57 | 0.5425 | 0.6139 |
| 58 | 0.5420 | 0.6134 |
| 59 | 0.5418 | 0.6131 |
| 60 | 0.5415 | 0.6128 |
| 61 | 0.5414 | 0.6127 |
| 62 | 0.5412 | 0.6123 |
| 63 | 0.5408 | 0.6119 |
| 64 | 0.5404 | 0.6116 |
| 65 | 0.5399 | 0.6109 |
| 66 | 0.5397 | 0.6108 |
| 67 | 0.5396 | 0.6107 |
| 68 | 0.5395 | 0.6105 |
| 69 | 0.5395 | 0.6104 |
| 70 | 0.5395 | 0.6104 |
| 71 | 0.5395 | 0.6104 |
| 72 | 0.5395 | 0.6103 |
| 73 | 0.5395 | 0.6103 |
| 74 | 0.5393 | 0.6101 |
| 75 | 0.5391 | 0.6099 |
| 76 | 0.5390 | 0.6098 |
| 77 | 0.5390 | 0.6099 |
| 78 | 0.5388 | 0.6096 |
| 79 | 0.5389 | 0.6097 |
| 80 | 0.5386 | 0.6095 |
| 81 | 0.5387 | 0.6096 |
| 82 | 0.5385 | 0.6094 |
| 83 | 0.5383 | 0.6092 |
| 84 | 0.5382 | 0.6091 |
| 85 | 0.5378 | 0.6087 |
| 86 | 0.5378 | 0.6087 |
| 87 | 0.5376 | 0.6084 |
| 88 | 0.5375 | 0.6084 |
| 89 | 0.5371 | 0.6080 |
| 90 | 0.5370 | 0.6079 |
| 91 | 0.5366 | 0.6075 |
| 92 | 0.5365 | 0.6074 |
| 93 | 0.5364 | 0.6073 |
| 94 | 0.5363 | 0.6072 |
| 95 | 0.5363 | 0.6072 |
| 96 | 0.5362 | 0.6072 |

| | | |
|----|--------|--------|
| 29 | 0.5980 | 0.6723 |
| 30 | 0.5981 | 0.6723 |
| 31 | 0.5976 | 0.6718 |
| 32 | 0.5971 | 0.6712 |
| 33 | 0.5967 | 0.6705 |
| 34 | 0.5964 | 0.6700 |
| 35 | 0.5961 | 0.6696 |
| 36 | 0.5961 | 0.6696 |
| 37 | 0.5959 | 0.6694 |
| 38 | 0.5958 | 0.6692 |
| 39 | 0.5952 | 0.6684 |
| 40 | 0.5946 | 0.6677 |
| 41 | 0.5947 | 0.6676 |
| 42 | 0.5946 | 0.6674 |
| 43 | 0.5944 | 0.6673 |
| 44 | 0.5942 | 0.6671 |
| 45 | 0.5940 | 0.6669 |
| 46 | 0.5941 | 0.6669 |
| 47 | 0.5940 | 0.6668 |
| 48 | 0.5934 | 0.6661 |
| 49 | 0.5931 | 0.6658 |
| 50 | 0.5926 | 0.6653 |
| 51 | 0.5925 | 0.6651 |
| 52 | 0.5925 | 0.6651 |
| 53 | 0.5925 | 0.6651 |
| 54 | 0.5924 | 0.6651 |
| 55 | 0.5923 | 0.6650 |
| 56 | 0.5922 | 0.6649 |
| 57 | 0.5921 | 0.6648 |
| 58 | 0.5920 | 0.6645 |
| 59 | 0.5917 | 0.6642 |
| 60 | 0.5916 | 0.6641 |
| 61 | 0.5917 | 0.6642 |
| 62 | 0.5917 | 0.6641 |
| 63 | 0.5915 | 0.6639 |
| 64 | 0.5913 | 0.6637 |
| 65 | 0.5914 | 0.6638 |
| 66 | 0.5913 | 0.6637 |
| 67 | 0.5913 | 0.6638 |
| 68 | 0.5913 | 0.6638 |
| 69 | 0.5913 | 0.6638 |
| 70 | 0.5912 | 0.6637 |
| 71 | 0.5910 | 0.6635 |
| 72 | 0.5909 | 0.6634 |
| 73 | 0.5909 | 0.6634 |
| 74 | 0.5909 | 0.6634 |
| 75 | 0.5910 | 0.6635 |
| 76 | 0.5910 | 0.6636 |
| 77 | 0.5911 | 0.6637 |
| 78 | 0.5911 | 0.6636 |
| 79 | 0.5912 | 0.6637 |
| 80 | 0.5912 | 0.6637 |
| 81 | 0.5912 | 0.6637 |
| 82 | 0.5912 | 0.6637 |
| 83 | 0.5913 | 0.6638 |
| 84 | 0.5913 | 0.6638 |
| 85 | 0.5913 | 0.6638 |

| | | |
|----|--------|--------|
| 86 | 0.5913 | 0.6638 |
| 87 | 0.5913 | 0.6638 |
| 88 | 0.5913 | 0.6638 |
| 89 | 0.5914 | 0.6639 |
| 90 | 0.5914 | 0.6639 |
| 91 | 0.5913 | 0.6638 |
| 92 | 0.5915 | 0.6640 |
| 93 | 0.5915 | 0.6640 |
| 94 | 0.5915 | 0.6640 |
| 95 | 0.5915 | 0.6640 |
| 96 | 0.5915 | 0.6640 |
| 97 | 0.5915 | 0.6639 |
| 98 | 0.5915 | 0.6639 |

ubuntu@ip-172-31-79-175:~/document-qa\$ python3.6 docqa/eval/ranked_scores.py output-confi.csv

Loading answers..

Computing ranks...

Rank answers_0_em answers_0_f1

| | | |
|----|--------|--------|
| 1 | 0.5565 | 0.6349 |
| 2 | 0.6091 | 0.6915 |
| 3 | 0.6180 | 0.6995 |
| 4 | 0.6215 | 0.7018 |
| 5 | 0.6225 | 0.7023 |
| 6 | 0.6195 | 0.6987 |
| 7 | 0.6193 | 0.6980 |
| 8 | 0.6176 | 0.6963 |
| 9 | 0.6155 | 0.6937 |
| 10 | 0.6140 | 0.6919 |
| 11 | 0.6125 | 0.6901 |
| 12 | 0.6103 | 0.6875 |
| 13 | 0.6094 | 0.6858 |
| 14 | 0.6080 | 0.6842 |
| 15 | 0.6067 | 0.6827 |
| 16 | 0.6056 | 0.6813 |
| 17 | 0.6047 | 0.6803 |
| 18 | 0.6041 | 0.6795 |
| 19 | 0.6039 | 0.6795 |
| 20 | 0.6031 | 0.6786 |
| 21 | 0.6027 | 0.6782 |
| 22 | 0.6024 | 0.6777 |
| 23 | 0.6012 | 0.6761 |
| 24 | 0.6007 | 0.6754 |
| 25 | 0.5994 | 0.6741 |
| 26 | 0.5987 | 0.6733 |
| 27 | 0.5981 | 0.6726 |
| 28 | 0.5981 | 0.6725 |
| 29 | 0.5980 | 0.6723 |
| 30 | 0.5981 | 0.6723 |
| 31 | 0.5976 | 0.6718 |
| 32 | 0.5971 | 0.6712 |
| 33 | 0.5967 | 0.6705 |
| 34 | 0.5964 | 0.6700 |
| 35 | 0.5961 | 0.6696 |
| 36 | 0.5961 | 0.6696 |
| 37 | 0.5959 | 0.6694 |
| 38 | 0.5958 | 0.6692 |

| | | |
|----|--------|--------|
| 39 | 0.5952 | 0.6684 |
| 40 | 0.5946 | 0.6677 |
| 41 | 0.5947 | 0.6676 |
| 42 | 0.5946 | 0.6674 |
| 43 | 0.5944 | 0.6673 |
| 44 | 0.5942 | 0.6671 |
| 45 | 0.5940 | 0.6669 |
| 46 | 0.5941 | 0.6669 |
| 47 | 0.5940 | 0.6668 |
| 48 | 0.5934 | 0.6661 |
| 49 | 0.5931 | 0.6658 |
| 50 | 0.5926 | 0.6653 |
| 51 | 0.5925 | 0.6651 |
| 52 | 0.5925 | 0.6651 |
| 53 | 0.5925 | 0.6651 |
| 54 | 0.5924 | 0.6651 |
| 55 | 0.5923 | 0.6650 |
| 56 | 0.5922 | 0.6649 |
| 57 | 0.5921 | 0.6648 |
| 58 | 0.5920 | 0.6645 |
| 59 | 0.5917 | 0.6642 |
| 60 | 0.5916 | 0.6641 |
| 61 | 0.5917 | 0.6642 |
| 62 | 0.5917 | 0.6641 |
| 63 | 0.5915 | 0.6639 |
| 64 | 0.5913 | 0.6637 |
| 65 | 0.5914 | 0.6638 |
| 66 | 0.5913 | 0.6637 |
| 67 | 0.5913 | 0.6638 |
| 68 | 0.5913 | 0.6638 |
| 69 | 0.5913 | 0.6638 |
| 70 | 0.5912 | 0.6637 |
| 71 | 0.5910 | 0.6635 |
| 72 | 0.5909 | 0.6634 |
| 73 | 0.5909 | 0.6634 |
| 74 | 0.5909 | 0.6634 |
| 75 | 0.5910 | 0.6635 |
| 76 | 0.5910 | 0.6636 |
| 77 | 0.5911 | 0.6637 |
| 78 | 0.5911 | 0.6636 |
| 79 | 0.5912 | 0.6637 |
| 80 | 0.5912 | 0.6637 |
| 81 | 0.5912 | 0.6637 |
| 82 | 0.5912 | 0.6637 |
| 83 | 0.5913 | 0.6638 |
| 84 | 0.5913 | 0.6638 |
| 85 | 0.5913 | 0.6638 |
| 86 | 0.5913 | 0.6638 |
| 87 | 0.5913 | 0.6638 |
| 88 | 0.5913 | 0.6638 |
| 89 | 0.5914 | 0.6639 |
| 90 | 0.5914 | 0.6639 |
| 91 | 0.5913 | 0.6638 |
| 92 | 0.5915 | 0.6640 |
| 93 | 0.5915 | 0.6640 |
| 94 | 0.5915 | 0.6640 |
| 95 | 0.5915 | 0.6640 |

| | | |
|----|--------|--------|
| 45 | 0.6232 | 0.7032 |
| 46 | 0.6233 | 0.7032 |
| 47 | 0.6233 | 0.7032 |
| 48 | 0.6230 | 0.7029 |
| 49 | 0.6227 | 0.7026 |
| 50 | 0.6227 | 0.7025 |
| 51 | 0.6225 | 0.7023 |
| 52 | 0.6226 | 0.7024 |
| 53 | 0.6226 | 0.7025 |
| 54 | 0.6226 | 0.7025 |
| 55 | 0.6227 | 0.7026 |
| 56 | 0.6227 | 0.7026 |
| 57 | 0.6228 | 0.7024 |
| 58 | 0.6227 | 0.7024 |
| 59 | 0.6226 | 0.7022 |
| 60 | 0.6226 | 0.7022 |
| 61 | 0.6226 | 0.7021 |
| 62 | 0.6226 | 0.7021 |
| 63 | 0.6226 | 0.7020 |
| 64 | 0.6225 | 0.7019 |
| 65 | 0.6226 | 0.7020 |
| 66 | 0.6226 | 0.7020 |
| 67 | 0.6227 | 0.7021 |
| 68 | 0.6226 | 0.7020 |
| 69 | 0.6226 | 0.7020 |
| 70 | 0.6226 | 0.7020 |
| 71 | 0.6226 | 0.7020 |
| 72 | 0.6226 | 0.7020 |
| 73 | 0.6226 | 0.7020 |
| 74 | 0.6226 | 0.7020 |
| 75 | 0.6228 | 0.7022 |
| 76 | 0.6227 | 0.7021 |
| 77 | 0.6228 | 0.7022 |
| 78 | 0.6228 | 0.7022 |
| 79 | 0.6228 | 0.7022 |
| 80 | 0.6226 | 0.7020 |
| 81 | 0.6226 | 0.7020 |
| 82 | 0.6225 | 0.7019 |
| 83 | 0.6226 | 0.7020 |
| 84 | 0.6226 | 0.7020 |
| 85 | 0.6226 | 0.7020 |
| 86 | 0.6226 | 0.7020 |
| 87 | 0.6226 | 0.7020 |
| 88 | 0.6226 | 0.7020 |
| 89 | 0.6226 | 0.7020 |
| 90 | 0.6226 | 0.7020 |
| 91 | 0.6226 | 0.7020 |
| 92 | 0.6227 | 0.7021 |
| 93 | 0.6227 | 0.7021 |
| 94 | 0.6227 | 0.7021 |
| 95 | 0.6227 | 0.7021 |
| 96 | 0.6227 | 0.7021 |
| 97 | 0.6227 | 0.7021 |
| 98 | 0.6227 | 0.7021 |

ubuntu@ip-172-31-79-175:~/document-qa\$ python3.6 docqa/eval/ranked_scores.py output-dr70.csv

Loading answers..

Computing ranks...

Rank answers_0_em answers_0_f1

| | | |
|----|--------|--------|
| 1 | 0.5496 | 0.6271 |
| 2 | 0.6074 | 0.6897 |
| 3 | 0.6185 | 0.7007 |
| 4 | 0.6237 | 0.7055 |
| 5 | 0.6275 | 0.7091 |
| 6 | 0.6287 | 0.7101 |
| 7 | 0.6299 | 0.7116 |
| 8 | 0.6299 | 0.7118 |
| 9 | 0.6293 | 0.7110 |
| 10 | 0.6293 | 0.7114 |
| 11 | 0.6293 | 0.7117 |
| 12 | 0.6289 | 0.7114 |
| 13 | 0.6286 | 0.7109 |
| 14 | 0.6278 | 0.7101 |
| 15 | 0.6277 | 0.7101 |
| 16 | 0.6272 | 0.7095 |
| 17 | 0.6267 | 0.7089 |
| 18 | 0.6264 | 0.7086 |
| 19 | 0.6267 | 0.7087 |
| 20 | 0.6263 | 0.7082 |
| 21 | 0.6264 | 0.7082 |
| 22 | 0.6259 | 0.7076 |
| 23 | 0.6255 | 0.7070 |
| 24 | 0.6251 | 0.7065 |
| 25 | 0.6251 | 0.7064 |
| 26 | 0.6250 | 0.7063 |
| 27 | 0.6249 | 0.7063 |
| 28 | 0.6251 | 0.7063 |
| 29 | 0.6249 | 0.7060 |
| 30 | 0.6251 | 0.7061 |
| 31 | 0.6251 | 0.7060 |
| 32 | 0.6251 | 0.7060 |
| 33 | 0.6248 | 0.7055 |
| 34 | 0.6246 | 0.7053 |
| 35 | 0.6244 | 0.7051 |
| 36 | 0.6244 | 0.7050 |
| 37 | 0.6238 | 0.7044 |
| 38 | 0.6237 | 0.7042 |
| 39 | 0.6238 | 0.7041 |
| 40 | 0.6235 | 0.7037 |
| 41 | 0.6236 | 0.7037 |
| 42 | 0.6235 | 0.7036 |
| 43 | 0.6235 | 0.7035 |
| 44 | 0.6234 | 0.7033 |
| 45 | 0.6232 | 0.7032 |
| 46 | 0.6233 | 0.7032 |
| 47 | 0.6233 | 0.7032 |
| 48 | 0.6230 | 0.7029 |
| 49 | 0.6227 | 0.7026 |
| 50 | 0.6227 | 0.7025 |
| 51 | 0.6225 | 0.7023 |
| 52 | 0.6226 | 0.7024 |
| 53 | 0.6226 | 0.7025 |
| 54 | 0.6226 | 0.7025 |
| 55 | 0.6227 | 0.7026 |
| 56 | 0.6227 | 0.7026 |

| | | |
|----|--------|--------|
| 57 | 0.6228 | 0.7024 |
| 58 | 0.6227 | 0.7024 |
| 59 | 0.6226 | 0.7022 |
| 60 | 0.6226 | 0.7022 |
| 61 | 0.6226 | 0.7021 |
| 62 | 0.6226 | 0.7021 |
| 63 | 0.6226 | 0.7020 |
| 64 | 0.6225 | 0.7019 |
| 65 | 0.6226 | 0.7020 |
| 66 | 0.6226 | 0.7020 |
| 67 | 0.6227 | 0.7021 |
| 68 | 0.6226 | 0.7020 |
| 69 | 0.6226 | 0.7020 |
| 70 | 0.6226 | 0.7020 |
| 71 | 0.6226 | 0.7020 |
| 72 | 0.6226 | 0.7020 |
| 73 | 0.6226 | 0.7020 |
| 74 | 0.6226 | 0.7020 |
| 75 | 0.6228 | 0.7022 |
| 76 | 0.6227 | 0.7021 |
| 77 | 0.6228 | 0.7022 |
| 78 | 0.6228 | 0.7022 |
| 79 | 0.6228 | 0.7022 |
| 80 | 0.6226 | 0.7020 |
| 81 | 0.6226 | 0.7020 |
| 82 | 0.6225 | 0.7019 |
| 83 | 0.6226 | 0.7020 |
| 84 | 0.6226 | 0.7020 |
| 85 | 0.6226 | 0.7020 |
| 86 | 0.6226 | 0.7020 |
| 87 | 0.6226 | 0.7020 |
| 88 | 0.6226 | 0.7020 |
| 89 | 0.6226 | 0.7020 |
| 90 | 0.6226 | 0.7020 |
| 91 | 0.6226 | 0.7020 |
| 92 | 0.6227 | 0.7021 |
| 93 | 0.6227 | 0.7021 |
| 94 | 0.6227 | 0.7021 |
| 95 | 0.6227 | 0.7021 |
| 96 | 0.6227 | 0.7021 |
| 97 | 0.6227 | 0.7021 |
| 98 | 0.6227 | 0.7021 |