

Homework 6

1 Introduction

1.1 Collaboration and Originality

1. Did you receive help of any kind from anyone in developing your software for this assignment (Yes or No)? It is not necessary to describe discussions with the instructor or TAs.

No

2. Did you give help of any kind to anyone in developing their software for this assignment (Yes or No)?

No

3. Did you examine anyone else's software for this assignment (Yes or No)? It is not necessary to mention software provided by the instructor.

No

4. Are you the author of every line of source code submitted for this assignment (Yes or No)? It is not necessary to mention software provided by the instructor.

Yes

5. Are you the author of every word of your report (Yes or No)?

Yes

1.2 Instructions

2 Experiment: Baselines

2.1 Experimental Results

	BM25 (Exp-2.1a)	Indri (Exp-2.1b)	LTR (Exp-2.1c)	DRMM (Exp-2.1d)
P@10	0.2800	0.2800	0.3760	0.2000
P@20	0.3020	0.2720	0.3860	0.2040
P@30	0.3093	0.2947	0.3640	0.2187
NDCG@10	0.1800	0.1861	0.2687	0.1493
NDCG@20	0.2144	0.1914	0.3080	0.1580
NDCG@30	0.2333	0.2168	0.3143	0.1797
MAP	0.2316	0.2144	0.2702	0.1303
Time	00:29	00:46	01:10	04:10

2.2 Parameters

```
indexPath=INPUT_DIR/index-cw09-nostem
queryFilePath=TEST_DIR/HW6-Exp-2.1a.qry
trecEvalOutputPath=OUTPUT_DIR/HW6-Exp-2.1a.teIn
trecEvalOutputLength=100
retrievalAlgorithm=BM25/Indri/letor
BM25:k1=1.2
BM25:b=0.75
BM25:k3=0
Indri:mu=1500
Indri:lambda=0.01
rerank=false/true
rerank:algorithm=drmm
rerank:maxInputRankingsLength=200
letor:trainingQueryFile=TEST_DIR/HW6-Exp-2.1a.trainQry
letor:trainingQrelsFile=TEST_DIR/HW6-Exp-2.1a.trainQrels
drmm:word2vecPath=INPUT_DIR/index-word2vec
drmm:numHistogramBins = 30
drmm:mlpLayers = 5,1
drmm:activations= tanh,tanh
drmm:numEpochs = 2500
drmm:numTrainingPairs = 5000
drmm:samplingStrategy = random
drmm:learningRate = 0.01
drmm:randomSeed=756756
drmm:shuffleSeed=2345
drmm:linearSeed=756756
```

2.3 Discussion

The DRMM model I implemented takes much longer time to run and worse performance comparing to the baseline. Comparing with later experiment, I found out maxInputRankingsLength has large influence on the performance. The DRMM can only rank files in a short ranking. The reason could be our implementation of DRMM is still very simple and not well trained.

3 Experiment: Effect of Training Strategies

3.1 Experimental results

Random Selection					
	2500 & 2500 (Exp- 3.1a)	1250 & 2500 (Exp-3.1b)	5000 & 2500 (Exp- 3.1c)	2500 & 5000 (Exp- 3.1d)	5000 & 5000 (Exp- 3.1e)
P@10	0.316	0.284	0.324	0.296	0.296
P@20	0.324	0.332	0.33	0.322	0.336
P@30	0.332	0.352	0.3307	0.328	0.336
NDCG@10	0.2046	0.205	0.2152	0.207	0.2197
NDCG@20	0.2397	0.2417	0.2393	0.2441	0.2541
NDCG@30	0.2662	0.2766	0.2631	0.2587	0.2751
MAP	0.2298	0.2303	0.2272	0.2208	0.2278
Time	03:50	02:35	06:05	04:59	06:18
Stratified Selection					
	2500 & 2500 (Exp-3.2a)	1250 & 2500 (Exp-3.2b)	5000 & 2500 (Exp-3.2c)	2500 & 5000 (Exp-3.2d)	5000 & 5000 (Exp-3.2e)
P@10	0.304	0.296	0.296	0.276	0.296
P@20	0.298	0.3	0.308	0.288	0.304
P@30	0.3227	0.316	0.3227	0.316	0.32
NDCG@10	0.2184	0.2099	0.2024	0.1903	0.1972
NDCG@20	0.2264	0.2284	0.2222	0.2109	0.2222
NDCG@30	0.255	0.2483	0.2514	0.2404	0.248
MAP	0.2292	0.2261	0.2252	0.2196	0.2253
Time	02:47	02:01	04:09	04:05	06:33

3.2 Parameters

```
indexPath=INPUT_DIR/index-cw09-nostem
queryFilePath=TEST_DIR/HW6-Exp-3.2a.qry
trecEvalOutputPath=OUTPUT_DIR/HW6-Exp-3.2a.teIn
trecEvalOutputLength=100
retrievalAlgorithm=BM25
BM25:k1=1.2
BM25:b=0.75
BM25:k3=0
Indri:mu=1500
Indri:lambda=0.01
rerank=TRUE
rerank:algorithm=drmm
rerank:maxInputRankingsLength=100
letor:trainingQueryFile=TEST_DIR/HW6-Exp-3.2a.trainQry
letor:trainingQrelsFile=TEST_DIR/HW6-Exp-3.2a.trainQrels
drmm:word2vecPath=INPUT_DIR/index-word2vec
drmm:numHistogramBins = 30
drmm:mlpLayers = 5,1
drmm:activations= tanh,tanh
drmm:numEpochs = 2500/5000
drmm:numTrainingPairs = 2500/1250/2500
drmm:samplingStrategy = stratified/random
drmm:learningRate = 0.01
drmm:randomSeed=756756
drmm:shuffleSeed=2345
drmm:linearSeed=756756
```

3.3 Discussion

The more pair it samples and more epoch it runs, the longer time it needs. And stratified strategy takes less time than random under same condition. Too less training pairs and too many training epochs could lead to overfit problem, while too many training pairs with too less training epochs could lead to underfit. Therefore, there will be a good combinational ratio between training pairs and training epochs. However, experiments show that more training pairs and more training epochs at the same time do not necessarily make the performance better.

4 Experiment: Effect of Histogram Bins

4.1 Experimental results

	10 (Exp-4.1a)	30 (Exp-4.1b)	5 (Exp-4.1c)	50 (Exp-4.1d)
P@10	0.208	0.316	0.376	0.332
P@20	0.24	0.324	0.364	0.324
P@30	0.2533	0.332	0.3547	0.3253
NDCG@10	0.1452	0.2046	0.2312	0.2401
NDCG@20	0.171	0.2397	0.271	0.2441
NDCG@30	0.1902	0.2662	0.2797	0.2559
MAP	0.1975	0.2298	0.2477	0.2231
Time	02:27	03:24	03:15	03:17

4.2 Parameters

indexPath=INPUT_DIR/index-cw09-nostem
queryFilePath=TEST_DIR/HW6-Exp-4.1a.qry
trecEvalOutputPath=OUTPUT_DIR/HW6-Exp-4.1a.teIn
trecEvalOutputLength=100
retrievalAlgorithm=BM25
BM25:k₁=1.2
BM25:b=0.75
BM25:k₃=0
Indri:mu=1500
Indri:lambda=0.01
rerank=TRUE
rerank:algorithm=drmm
rerank:maxInputRankingsLength=100
letor:trainingQueryFile=TEST_DIR/HW6-Exp-4.1a.trainQry
letor:trainingQrelsFile=TEST_DIR/HW6-Exp-4.1a.trainQrels
drmm:word2vecPath=INPUT_DIR/index-word2vec
drmm:numHistogramBins = 10/30/5/50
drmm:mlpLayers = 5,1
drmm:activations= tanh,tanh
drmm:numEpochs = 2500
drmm:numTrainingPairs = 2500
drmm:samplingStrategy = random
drmm:learningRate = 0.01
drmm:randomSeed=756756
drmm:shuffleSeed=2345
drmm:linearSeed=756756

4.3 Discussion

More bins do not necessarily make the performance better. Bin with 5 has a better score on all metrics, $P@10$, $P@20$, $P@30$, $NDCG@10$, $NDCG@20$, $NDCG@30$ and MAP. And the change of these metrics are mostly consistent with respect to the bin size. Bin 5 has the highest score and bin 10 has the lowest.

5 Experiment: The effect of the Feedforward Network

5.1 Experimental results

	1/tanh (Exp-5.1a)	10,5,1/tanh,tanh (Exp-5.1b)	10,5,1/relu,relu,relu (Exp-5.1c)	10,10,1/tanh,tanh,tanh (Exp-5.1d)
P@10	0.3	0.252	0.348	0.24
P@20	0.31	0.26	0.356	0.256
P@30	0.3147	0.2667	0.3413	0.2667
NDCG@10	0.1798	0.1484	0.2503	0.139
NDCG@20	0.2089	0.163	0.2724	0.16
NDCG@30	0.2294	0.1891	0.2851	0.1877
MAP	0.2235	0.1994	0.2469	0.1983
Time	02:18	03:48	03:13	04:10

5.2 Parameters

indexPath=INPUT_DIR/index-cw09-nostem
queryFilePath=TEST_DIR/HW6-Exp-5.1a.qry
trecEvalOutputPath=OUTPUT_DIR/HW6-Exp-5.1a.teIn
trecEvalOutputLength=100
retrievalAlgorithm=BM25
BM25:k₁=1.2
BM25:b=0.75
BM25:k₃=0
Indri:mu=1500
Indri:lambda=0.01
rerank=TRUE
rerank:algorithm=drmm
rerank:maxInputRankingsLength=100
letor:trainingQueryFile=TEST_DIR/HW6-Exp-5.1a.trainQry
letor:trainingQrelsFile=TEST_DIR/HW6-Exp-5.1a.trainQrels
drmm:word2vecPath=INPUT_DIR/index-word2vec
drmm:numHistogramBins = 5
drmm:numEpochs = 2500
drmm:numTrainingPairs = 2500
drmm:samplingStrategy = random
drmm:learningRate = 0.01
drmm:randomSeed=756756
drmm:shuffleSeed=2345
drmm:linearSeed=756756

5.4 Discussion

The network configuration has large influence on the performance. Firstly, the activation has large influence. Exp-c and Exp-b have same number of layer and units. But the relu network is much better than the tanh network. Larger network not necessarily make the performance better. Exp-a, Exp-b and Exp-d have same activation function. But the larger the network it is the performance goes worse. The reason could be that the model is not well trained enough due to the limited data.