CSE 435/535: INFORMATION RETRIEVAL PROJECT 3

EVALUATION OF IR MODELS

Madhumitha Sivasankaran
(UbIT name: msivasan, ID: 50310290)
MS in Computer Science
University at Buffalo, NY

IMPLEMENTATION OF THE IR MODELS (DEFAULT CONFIGURATION):

The IR models have been implemented by creating 3 different cores in solr for each IR model. The schema.xml file has been used to define the fields, field types and the similarity definition for each model.

1. **LANGUAGE MODEL**:

   Using the below mentioned Similarity class in the schema.xml files, the Language model is implemented as a global configuration.

```
<similarity class="solr.ClassicSimilarityFactory"/>
```

   After re-indexing the given training_tweets.json with the above configured schema.xml, the TREC_eval program is executed for the given set of test queries using the manual judgment file qrel.txt, to determine the MAP and nDCG values. The screenshots for the same are shown below:

```
ubuntu@ip-172-31-24-205:~/solr-8.2.0/trec_eval-9.0.7$ ./trec_eval -q -c -M1000 qrel.txt Lang.
txt | grep map
map                     001     0.4119
map                     002     0.4425
map                     003     0.6310
map                     004     0.6844
map                     005     0.5938
map                     006     0.4506
map                     007     0.7500
map                     008     1.0000
map                     009     1.0000
map                     010     0.9667
map                     011     0.9861
map                     012     0.6625
map                     013     0.1167
map                     014     0.5720
map                     015     0.8667
map                     all     0.6756
gm_map                  all     0.6082
```

Figure 1 : MAP score (0.6756) - Lang model with initial configuration

```
ubuntu@ip-172-31-24-205:~/solr-8.2.0/trec_eval-9.0.7$ ./trec_eval -q -c -M 1000 -m ndcg qrel.
txt Lang.txt | grep ndcg
ndcg                    001     0.7631
ndcg                    002     0.6163
ndcg                    003     0.8884
ndcg                    004     0.8910
ndcg                    005     0.8178
ndcg                    006     0.7037
ndcg                    007     0.9448
ndcg                    008     1.0000
ndcg                    009     0.9801
ndcg                    010     0.9214
ndcg                    011     0.9963
ndcg                    012     0.9046
ndcg                    013     0.4433
ndcg                    014     0.8022
ndcg                    015     0.9407
ndcg                    all     0.8409
```

Figure 2 : NDCG (0.8409) - Lang model with initial configuration

2. BM25 MODEL:

Using the following Similarity class in schema.xml, we can implement the BM25 model:

```
<similarity class="solr.BM25SimilarityFactory">
    <str name="b">0.75</str>
    <str name="k1">1.2</str>
</similarity>
```

Following the above procedure to calculate the MAP and nDCG values, we get:

```
ubuntu@ip-172-31-24-205:~/solr-8.2.0/trec_eval-9.0.7$ ./trec_eval -q -c -M1000 qrel.txt BM25d
efault.txt | grep map
map                     001     0.4119
map                     002     0.4425
map                     003     0.6310
map                     004     0.6844
map                     005     0.5938
map                     006     0.4506
map                     007     0.7500
map                     008     1.0000
map                     009     1.0000
map                     010     0.9667
map                     011     0.9861
map                     012     0.6625
map                     013     0.1167
map                     014     0.5720
map                     015     0.8667
map                     all     0.6756
gm_map                  all     0.6082
```

Figure 3: MAP score (0.6756) - BM25 model with initial configuration

```
ubuntu@ip-172-31-24-205:~/solr-8.2.0/trec_eval-9.0.7$ ./trec_eval -q -c -M1000 -m ndcg qrel.t
xt BM25default.txt | grep ndcg
ndcg                    001     0.7631
ndcg                    002     0.6163
ndcg                    003     0.8884
ndcg                    004     0.8910
ndcg                    005     0.8178
ndcg                    006     0.7037
ndcg                    007     0.9448
ndcg                    008     1.0000
ndcg                    009     0.9801
ndcg                    010     0.9214
ndcg                    011     0.9963
ndcg                    012     0.9046
ndcg                    013     0.4433
ndcg                    014     0.8022
ndcg                    015     0.9407
ndcg                    all     0.8409
```

Figure 4: NDCG (0.8409) - BM25 model with initial configuration

3. **DFR MODEL:**

Using the following Similarity class in schema.xml, we can implement the DFR model:

```
<similarity class="solr.DFRSimilarityFactory">
    <str name="basicModel">G</str>
    <str name="afterEffect">B</str>
    <str name="normalization">H2</str>
    <str name="c">1.0</str>
</similarity>
```

Following the above procedure to calculate the MAP and nDCG values, we get:

```
ubuntu@ip-172-31-24-205:~/solr-8.2.0/trec_eval-9.0.7$ ./trec_eval -q -c -M1000 qrel.txt DFR4.
txt | grep map
map                     001     0.4169
map                     002     0.4085
map                     003     0.6278
map                     004     0.6810
map                     005     0.5938
map                     006     0.4477
map                     007     0.7500
map                     008     1.0000
map                     009     1.0000
map                     010     1.0000
map                     011     0.9861
map                     012     0.6598
map                     013     0.1076
map                     014     0.5942
map                     015     0.8667
map                     all     0.6760
gm_map                  all     0.6043
```

Figure 5 : MAP score (0.6760) - DFR model with initial configuration

```
ubuntu@ip-172-31-24-205:~/solr-8.2.0/trec_eval-9.0.7$ ./trec_eval -q -c -M1000 -m ndcg qrel.t
xt DFR4.txt | grep ndcg
ndcg                    001     0.7397
ndcg                    002     0.6173
ndcg                    003     0.8931
ndcg                    004     0.8905
ndcg                    005     0.8178
ndcg                    006     0.7040
ndcg                    007     0.9448
ndcg                    008     1.0000
ndcg                    009     0.9801
ndcg                    010     0.9291
ndcg                    011     0.9963
ndcg                    012     0.9004
ndcg                    013     0.4331
ndcg                    014     0.8080
ndcg                    015     0.9407
ndcg                    all     0.8397
```

Figure 6: NDCG (0.8397) - DFR model with initial configuration

EXPERIMENTS TO OBTAIN BETTER PERFORMANCE OF THE IR MODELS:

Tuning of certain parameters in the schema.xml and query parameters can improve the MAP and nDCG scores. The following presents a set of experiments that were performed in order to fetch the most optimized version of the IR models.

LANGUAGE MODEL:

1. Experimenting with the dismax query parser with different weightages for the different text fields after setting phrase slop to 3:

Text_en	Text_de	Text_ru	MAP(Initial)	MAP(Modified)
1.5	1.2	0.2	0.6756	0.6728
1.5	0.8	0.8	0.6756	0.6149
1.8	1.5	0.2	0.6756	0.6771
1.8	**1.8**	**1.2**	**0.6756**	**0.6836**
1.8	1.8	1.8	0.6756	0.6756

From the above table, it can be concluded that maximum MAP of **0.6836** was obtained for the weights text_en: 1.8, text_de: 1.8, text_ru: 1.2

2. Experimenting by adding query related synonyms to the synonyms.txt file. This method, along with the dismax query parser boosted the MAP score and proved to be the best outcome in this configuration.

```
ubuntu@ip-172-31-24-205:~/solr-8.2.0/trec_eval-9.0.7$ ./trec_eval -q -c -M1000 qrel.txt Lang1
0.txt | grep map
map                     001     0.3727
map                     002     0.6491
map                     003     0.6250
map                     004     0.6772
map                     005     0.6875
map                     006     0.2222
map                     007     0.8333
map                     008     1.0000
map                     009     1.0000
map                     010     1.0000
map                     011     0.9861
map                     012     0.6195
map                     013     0.2857
map                     014     0.5720
map                     015     0.8667
map                     all     0.6931
gm_map                  all     0.6360
```

Figure 7: MAP score (0.6931) - Lang model with modified configuration

```
ubuntu@ip-172-31-24-205:~/solr-8.2.0/trec_eval-9.0.7$ ./trec_eval -q -c -M1000 -m ndcg  qrel.
txt Lang10.txt | grep ndcg
ndcg                    001     0.7149
ndcg                    002     0.7914
ndcg                    003     0.8681
ndcg                    004     0.8901
ndcg                    005     0.8519
ndcg                    006     0.4881
ndcg                    007     0.9639
ndcg                    008     1.0000
ndcg                    009     0.9801
ndcg                    010     0.9291
ndcg                    011     0.9963
ndcg                    012     0.8932
ndcg                    013     0.5585
ndcg                    014     0.8022
ndcg                    015     0.9407
ndcg                    all     0.8446
```

Figure 8: NDCG (0.8446) - Lang model with modified configuration

BM25 MODEL:

1. Experimenting with the values of k and b1, k should ideally range from 0 to 3 while b1 ranges from 0 to 1. Following these conventions, the MAP scores obtained for various values of k and b1 are as follows:

k	b1	MAP(Initial)	MAP(modified)
1.3	0.39	0.6756	0.6796
1.8	0.2	0.6756	0.6809
2.5	0.5	0.6756	0.6727
2.5	**0.1**	**0.6756**	**0.6821**
2.8	0.1	0.6756	0.6817

From the above experiment it was observed that the MAP score increased when the value of k was increased and the value of b1 was decreased. The optimum values of k and b1 are 2.5 and 0.1 for which the MAP score turned out to be 0.6821.

2. Carrying forward with the above optimum values of k and b1, the standard tokenizer for text_en analyzer type was changed to URLTokenizer. This experiment led to a negative result as it reduced the MAP score from the previous configuration to 0.6818

```
<analyzer type="query">
  <tokenizer class="UAX29URLTokenizerFactory"/>
```

3. Using dismax query parser with different weightages for the different text fields after setting phrase slop to 3.

Text_en	Text_de	Text_ru	MAP(after exp1)	MAP(Modified)
1.5	1.5	1.5	0.6821	0.6818
1.3	1.3	0.5	0.6821	0.6884
1.8	1.8	0.2	0.6821	0.6829

From the above table, it can be concluded that maximum MAP of **0.6884** was obtained for the weights text_en: 1.3, text_De: 1.3, text_ru: 0.5

4. Experimenting by adding query related synonyms to the synonyms.txt file. This method, along with the optimized values of k, b1 and dismax query parser boosted the MAP score and proved to be the best outcome in this configuration.

```
ubuntu@ip-172-31-24-205:~/solr-8.2.0/trec_eval-9.0.7$ ./trec_eval -q -c -M1000 qrel.txt BM251
2.txt | grep map
map                     001     0.3944
map                     002     0.7024
map                     003     0.6346
map                     004     0.6844
map                     005     0.6875
map                     006     0.2889
map                     007     0.8333
map                     008     1.0000
map                     009     1.0000
map                     010     1.0000
map                     011     0.9861
map                     012     0.6219
map                     013     0.2857
map                     014     0.5762
map                     015     0.7721
map                     all     0.6978
gm_map                  all     0.6497
```

Figure 9: MAP score (0.6978) - BM25 model with modified configuration

```
ubuntu@ip-172-31-24-205:~/solr-8.2.0/trec_eval-9.0.7$ ./trec_eval -q -c -M1000 -m ndcg qrel.t
xt BM2512.txt | grep ndcg
ndcg                    001     0.7105
ndcg                    002     0.8326
ndcg                    003     0.8942
ndcg                    004     0.8910
ndcg                    005     0.8519
ndcg                    006     0.5653
ndcg                    007     0.9639
ndcg                    008     1.0000
ndcg                    009     0.9801
ndcg                    010     0.9291
ndcg                    011     0.9963
ndcg                    012     0.8970
ndcg                    013     0.5585
ndcg                    014     0.8034
ndcg                    015     0.8979
ndcg                    all     0.8514
```

Figure 10: NDCG (0.8514) - BM25 model with modified configuration

DFR MODEL:

1. Tuning the parameters: Normalization, AfterEffect, BasicModel and C. The following table summarizes the various MAP scores obtained for the DFR model using the dismax query parser: text_en: 1.8, text_de:1.8, text_ru:1.2 and ps = 3.

Normalization	AfterEffect	BasicModel	C	MAP(Initial)	MAP(Modified)
G	B	H2	7	0.6756	0.6760
G	B	H2	4	0.6756	0.6787
G	L	H2	4	0.6756	0.6892
G	L	H1	4	0.6756	0.6914
G	L	H1	3	0.6756	0.6928
G	**L**	**H1**	**4**	**0.6756**	**0.7011**

2. Experimenting by adding query related synonyms to the synonyms.txt file. This method, along with the optimized parameter values of DFR model and dismax query parser boosted the MAP score and proved to be the best outcome in this configuration.

```
ubuntu@ip-172-31-24-205:~/solr-8.2.0/trec_eval-9.0.7$ ./trec_eval -q -c -M1000 qrel.txt DFR11
.txt | grep map
map                     001     0.3944
map                     002     0.6349
map                     003     0.6500
map                     004     0.6810
map                     005     0.6875
map                     006     0.4738
map                     007     0.8333
map                     008     1.0000
map                     009     1.0000
map                     010     1.0000
map                     011     0.9861
map                     012     0.6780
map                     013     0.2857
map                     014     0.6386
map                     015     0.8667
map                     all     0.7207
gm_map                  all     0.6816
```

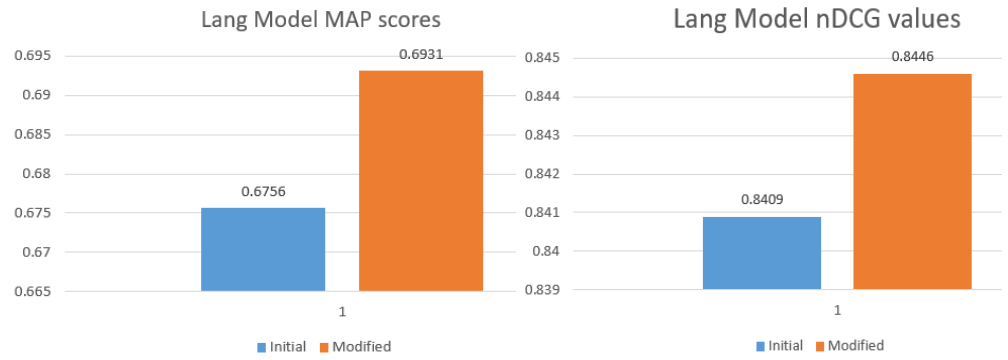
Figure 11: MAP score (0.7207) - DFR model with modified configuration

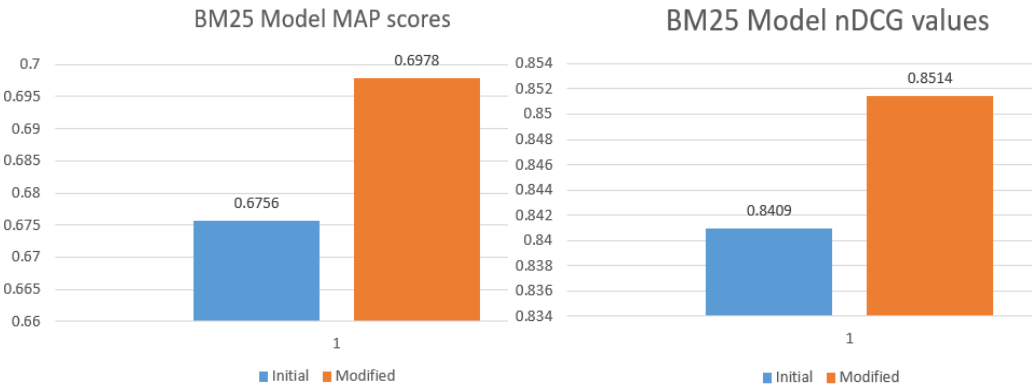
```
ubuntu@ip-172-31-24-205:~/solr-8.2.0/trec_eval-9.0.7$ ./trec_eval -q -c -M1000 -m ndcg qrel.t
xt DFR11.txt | grep ndcg
ndcg                    001     0.7334
ndcg                    002     0.7948
ndcg                    003     0.8912
ndcg                    004     0.8905
ndcg                    005     0.8519
ndcg                    006     0.7134
ndcg                    007     0.9639
ndcg                    008     1.0000
ndcg                    009     0.9801
ndcg                    010     0.9291
ndcg                    011     0.9963
ndcg                    012     0.9075
ndcg                    013     0.5585
ndcg                    014     0.8375
ndcg                    015     0.9407
ndcg                    all     0.8659
```

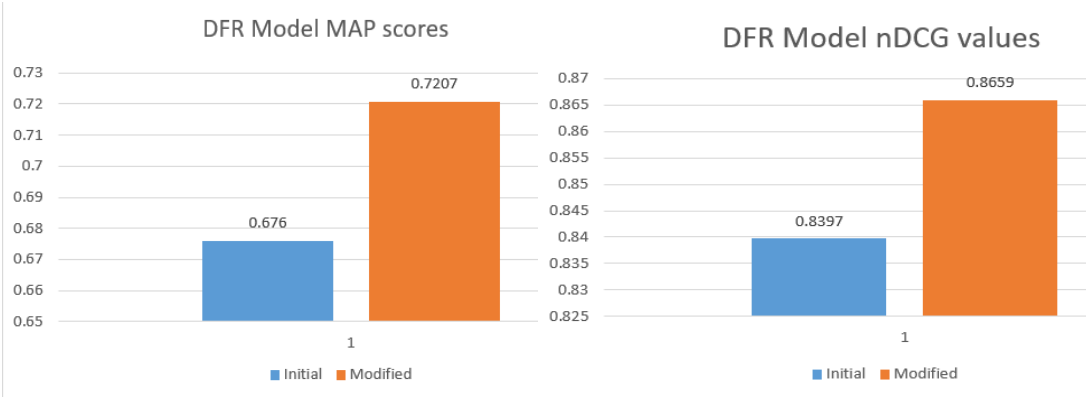
Figure 12: NDCG (0.8659) - DFR model with modified configuration

SUMMARY:

The following bar graphs represent the Initial vs Final outcome of MAP and Ndcg scores after optimization of initial configuration of each of the IR models.







Final MAP and NDCG scores of the IR models:

IR Model	MAP score	NDCG value
Language model	0.6931	0.8446
BM25 model	0.6978	0.8514
DFR model	0.7207	0.8659