Evaluation of IR Models

Priya Rao

UBIT Name: prao4

Person ID: 50321961

Introduction

The goal of this project is to implement various IR models, evaluate the IR system and improve the search results based on the understanding of the models, the implementation and the evaluation. Given is twitter data in three languages - English, German and Russian, 15 sample queries and the corresponding relevance judgements. These are indexed using Solr and the following three IR models are implemented: (i) Language Model, (ii) BM25 and (iii) Divergence from Randomness (DFR) Model. The results from these three sets will be evaluated using the Trec_eval program. Based on the evaluation results, the performance in terms of Mean Average Precision (MAP) is tweaked to get the best results.

Implementation

The three models in solr have been implemented using the similarity class which needs to be added into the schema.xml. Since in this project we have 3 models to implement, we would be building 3 schema files which contain the similarity class implementation of BM25, LM and DFR as shown (below are the implementations with the default values):

```
<schema name="default-config" version="1.6">
  <uniqueKey>id</uniqueKey>
  <similarity class="solr.BM25SimilarityFactory">
  <float name="b">0.75</float>
  <float name="k1">1.2</float>
  </similarity>
```

```
<schema name="default-config" version="1.6">
  <uniqueKey>id</uniqueKey>
  <similarity class="solr.LMDirichletSimilarityFactory">
  </similarity>
```

```
<schema name="default-config" version="1.6">
  <uniqueKey>id</uniqueKey>
  <similarity class="solr.DFRSimilarityFactory">
    <str name="normalization">H2</str>
    <str name="afterEffect">B</str>
    <str name="basicModel">G</str>
  </similarity>
```

The parameters of each of the schemas must be tweaked in order to check which parameters would achieve the best similarity scores.

Using the default parameters and having used no query parsers or any form of optimization done, the below values were obtained for BM25, LM and DFR:

```
runid                   all     BM25
num_q                   all     15
num_ret                 all     280
num_rel                 all     225
num_rel_ret             all     129
map                     all     0.6985
gm_map                  all     0.6302
```

```
runid                   all     DFR
num_q                   all     15
num_ret                 all     280
num_rel                 all     225
num_rel_ret             all     130
map                     all     0.7055
gm_map                  all     0.6382
```

```
runid                   all     LM
num_q                   all     15
num_ret                 all     280
num_rel                 all     225
num_rel_ret             all     119
map                     all     0.6299
gm_map                  all     0.5489
```

Strategies

Upon tweaking the parameters of the BM25 model, i.e., decreasing both b and k1 values by 0.1 (0.6 and 1.1), we get:

```
runid                   all     BM25
num_q                   all     15
num_ret                 all     280
num_rel                 all     225
num_rel_ret             all     130
map                     all     0.6955
gm_map                  all     0.6285
```

This shows that decreasing the b and k1 values in the naïve model gives a decrease in the map score. Increasing the b and k1 values (0.9 and 1.4) in the naïve model increases the map score.

```
runid                   all     BM25
num_q                   all     15
num_ret                 all     280
num_rel                 all     225
num_rel_ret             all     129
map                     all     0.7004
gm_map                  all     0.6329
```

Upon further increasing the value of both b and k1 (1.0 and 1.5) in the naïve model, there is a decrease in the overall map score:

```
runid                   all     BM25
num_q                   all     15
num_ret                 all     280
num_rel                 all     225
num_rel_ret             all     130
map                     all     0.6995
gm_map                  all     0.6340
```

For k1 = 1.5 and b = 0.75, we get the values as:

```
runid                   all     BM25
num_q                   all     15
num_ret                 all     280
num_rel                 all     225
num_rel_ret             all     141
map                     all     0.7019
gm_map                  all     0.6686
```

Which is so far the ideal score obtained.

Upon adding a custom synonyms file, which allows for all those words not accounted by the query but are accounted for in the document which convey a similar meaning to that of the querying term, the BM25, LM and DFR models produce the below map scores:

```
runid                   all     BM25
num_q                   all     15
num_ret                 all     280
num_rel                 all     225
num_rel_ret             all     138
map                     all     0.6905
```

```
runid                   all     LM
num_q                   all     15
num_ret                 all     280
num_rel                 all     225
num_rel_ret             all     125
map                     all     0.6325
gm_map                  all     0.5655
```

```
runid                   all     DFR
num_q                   all     15
num_ret                 all     280
num_rel                 all     225
num_rel_ret             all     141
map                     all     0.7117
gm_map                  all     0.6784
```

Upon tweaking the parameters of LM (mu = 1800), we can see a maximum map score of:

```
runid                   all     LM
num_q                   all     15
num_ret                 all     280
num_rel                 all     225
num_rel_ret             all     125
map                     all     0.6346
gm_map                  all     0.5686
```

Observations

For BM25

K1	b	MAP score
1.2 (default)	0.75 (default)	0.6985
1.1	0.6	0.6955
1.4	0.9	0.7004
1.5	1.0	0.6995
1.5	0.75	0.7019

For DFR

Basicmodel	Aftereffect	Normalization	MAP score
G	B	H2	0.7117

For LM

Mu	MAP Score
2000 (default)	0.6325
1800	0.6346
1500	0.6286

Results

The best model evaluated among the three models is DFR with a MAP score of 0.7117, followed by BM25 with a MAP score of 0.7019 and LM with a MAP score of 0.6346.