Evaluation of Various Probabilistic IR Models

Vineeth Thayanithi

November 2019

1 Introduction

Probabilistic IR models are a set of models that uses ranking functions to rank documents that are present in the order of their corresponding relevance with the query. For this project, we would be implementing the models (i) BM25 (ii) Language (iii) Divergence from Randomness on solr 8.2.0. Once we have implemented these models, we would evaluate the models against various parameters.

2 Dataset

The data set that would be posted to solr for indexing are a set of processed tweets. This dataset contains approximately about 3500 documents in total. The tweets extracted maybe in one of three languages Russian, English or German.

3 Implementation

3.1 General Implementation

The solr instance instance is run on amazon aws instance running on ubuntu. Once the solr instance is up and running on port 8984, we move on to to create 3 separate cores for the 3 models stated previously. We use the similarity class identifier for implementing the models.

3.2 Implementing BM25 Model

The BM25 Model is the default IR model of solr. The BM25 model can be implemented by using the BM25SimilarityFactory. The BM25SimilarityFactory class takes two parameters b and k1. b determines the normalization factor of tf values by document length and k1 specifies the non-linear term frequency. The default values of b and k1 are 0.75 and 1.2 respectively.

3.3 Implementing Language Model

The language model can be implemented by using the LMDirichiletSimilarity-Factory. The Language model works by assigning a negative score to documents that contain a term but for a lesser number of times than the model had predicted. mu is the parameter that this similarity class uses and its default value is 2000.

3.4 Implmenting Divergence From Randomness Model

The Divergence from randomness model can be implemented by using the DFR-SimilarityFactory. The DFR scoring component is based on 3 factors, the basic model, the after effect and the normalization type. The basic model G uses the geometric approximation of Bose einstein distribution, The after effect B is for first normalization that is the ratio of two Bernoulli processes and finally H2 is the second normalization which normalizes term frequency density inversely related to length.

4 Evaluation

Once these models are up and running we run our customized python script against these models for obtaining the scores for 20 best matches of each query along with their ranks. This is stored in text files for each query and text files for each model is tested against the trec eval program using the relevance relation that has been computed already. From the default parameter and query parser, the map values for each model were recorded to be as,

- BM25 - 0.3750
- Language - 0.3377
- DFR - 0.3695

Using the trec eval results we then move forward for tweaking the models.

5 Tweaking

5.1 Query Parser

Query parsers are specified in solrconfig.xml. The default query parser of solr has very strict syntax checks during query processing which could lead to a lot of loss in number of relevant documents. Thus we use the extended dismax query processor to our models. The query fields of the query parser uses the fields text_en, text_ru, text_de and tweet_hashtag. The three text language fields are boosted by a factor of 4 whereas the tweet_hashtag field is boosted by a factor of 5 since the hashtags give a gist of the entire content. The qs parameter is set to 15.

5.2 Tuning BM25 Model

The parameter b can take values from 0 to 1 and the parameter k1 can be between 0 and 3. on varying the parameters the below is the observation that was recorded.

S.No	K1	b	MAP
1	0.4	0.9	0.6990
2	0.4	1.0	0.6992
3	0.4	1.1	0.6997
4	0.4	1.5	0.7003
5	0.4	1.6	0.7003
5	0.4	2.0	0.6976

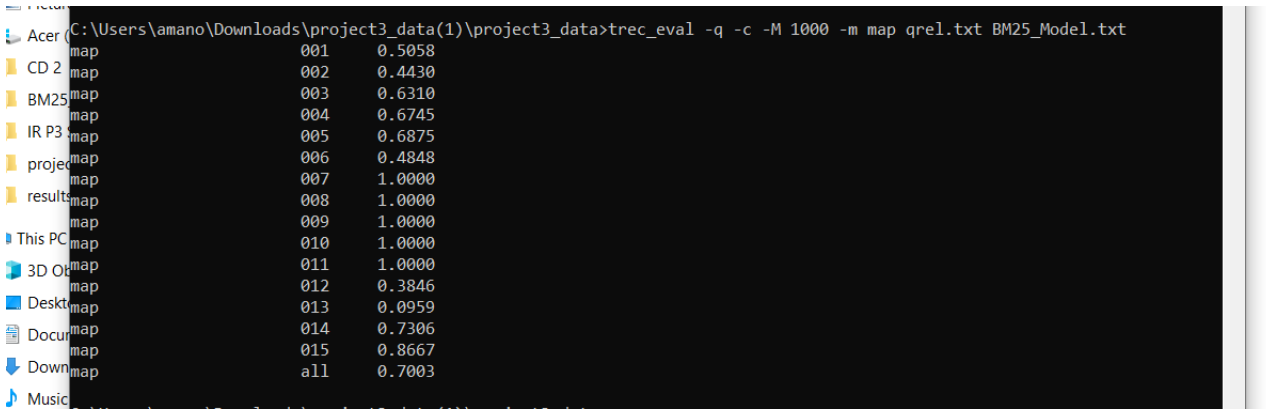

Figure 1: BM25 Model MAP Values

5.3 Optimizing Language Model

The initial map of the language model with edismax parser was found to be as 0.6241 and the map had begun to alter after changing the mu values and the variation is as below.

S.No	mu	MAP
1	2000	0.6241
2	500	0.6661
3	100	0.6950
4	50	0.6984
4	25	0.7048
5	10	0.7048

```
C:\Users\amano\Downloads\project3_data(1)\project3_data>trec_eval -q -c -M 1000 -m map qrel.txt Language_Model.txt
map                     001     0.4926
map                     002     0.5095
map                     003     0.6500
map                     004     0.6810
map                     005     0.6875
map                     006     0.4848
map                     007     1.0000
map                     008     1.0000
map                     009     1.0000
map                     010     1.0000
map                     011     1.0000
map                     012     0.3846
map                     013     0.1023
map                     014     0.7127
map                     015     0.8667
map                     all     0.7048
```

Figure 2: Language Model MAP Values

5.4 Optimizing DFR Model

The parameters for dfr model are predefined and no other optimizations were made apart from using edismax query parser for DFR Model.

```
C:\Users\amano\Downloads\project3_data(1)\project3_data>trec_eval -q -c -M 1000 -m map qrel.txt DFR_Model.txt
map                     001     0.3924
map                     002     0.4204
map                     003     0.6346
map                     004     0.6470
map                     005     0.6500
map                     006     0.4774
map                     007     1.0000
map                     008     1.0000
map                     009     1.0000
map                     010     1.0000
map                     011     1.0000
map                     012     0.3462
map                     013     0.0913
map                     014     0.7306
map                     015     0.8667
map                     all     0.6838
```

Figure 3: DFR Model MAP Values

6 Conclusion

All of the models have been implemented on solr and their MAP values are optimized. Out of the three models, the Language Model has the highest MAP value of 0.7048. The BM25 model ended up with a 0.7003 MAP while DFR came with a 0.6838 MAP. The study thus shows how different models react in different situtions.