

**INTRODUCTION TO INFORMATION
RETRIEVAL**

CSE 535

FALL-2019

PROJECT-3 REPORT

EVALUATION OF IR MODELS

SUBMITTED BY:

ROHITH KUMAR GADALAY(50314899)(rohithku)

EVALUATION OF INFORMATION RETRIEVAL MODELS

1. INTRODUCTION

This project deals with the implementation of IR models such as LM model, BM25 model, DFR models based on Solr using twitter data and the results are evaluated using TREC_eval program.

We are given 15 training queries and 10 testing queries in languages – German, English and Russian. The key concept of this project is to improve the performance of IR system by considering primarily MAP (mean average precision) as an evaluation measure.

2. OVERALL METHODOLOGY

Since the intention of this project is to improve the MAP (mean average precision) value, the experimentation is done on the three models – LM, BM25 and DFR under different set of circumstances and the mean average precision values for each of these are collected. Initially, the MAP values are determined for the three similarity models without using any parsers. Then these MAP values are compared with implementations using query parsers like dismax.

3. EXPERIMENTATION:

I have created 3 cores – one for each model – LM(Language Model), BM25, DFR(Divergence From Randomness) by modifying the schema.xml of each model. The following similarity classes have been used for each model:

i) LM

In Language Model(LM), a document is good match to a query if the document model is likely to generate a query, which will in turn happen if the document contains the query words often. This approach thus provides a different realization of some of the basic ideas for document ranking. In Solr, the corresponding similarity class is solr.LMDirichletSimilarityFactory.

The syntax in schema.xml is:

```
</analyzer>
<similarity class="solr.LMDirichletSimilarityFactory">
  <float name="mu">0.2</float>
</similarity>
```

ii) BM25

BM25 model is a probabilistic information retrieval model which was originally designed for short-length documents. In Solr, the similarity class for this is `solr.BM25SimilarityFactory`.

The syntax in `schema.xml` is:

```
<similarity class="solr.BM25SimilarityFactory">
  <float name="k1">0.2</float>
  <float name="b">0.6</float>
</similarity>
```

iii) DFR

In Divergence from randomness model, the term-weight is inversely related to the probability of term-frequency within the document obtained by a model of randomness.

In Solr, the similarity class for this is given by `solr.DFRSimilarityFactory`.

The syntax in `schema.xml` is:

```
<similarity class="solr.DFRSimilarityFactory">
  <str name="basicModel">I(ne)</str>
  <str name="afterEffect">L</str>
  <str name="normalization">H2</str>
  <float name="c">10</float>
</similarity>
```

Using the default settings and without any parser, I have obtained quite low values.

The MAP values were found out to be as follows for the default settings:

i) LM - 0.2521-default $\mu=2000$

ii) BM25 – 0.2451 - default $k_1 = 1.2$ and $b = 0.75$

iii) DFR – 0.2466 – given defaults – H2 normalization, Basic model G and Bernoulli

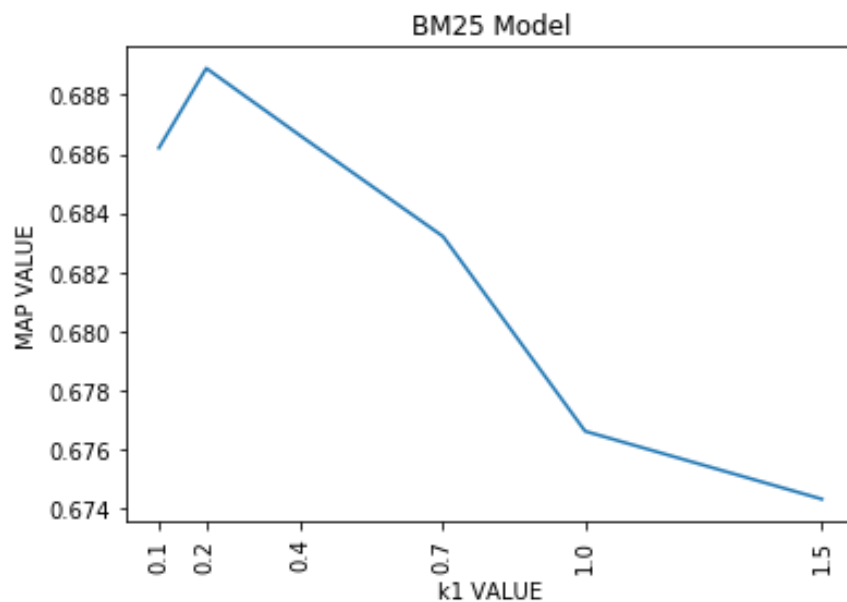
A python script is developed which parses the queries text file one by one and returns the query results into a new text file for each of the models by running the query url such that 3 text files are produced corresponding to each set of the core (LM, BM25, DFR) and query list. The models are run against `queries.txt` and `test_queries.txt` file. The retrieved output and given `qrel.txt` files are evaluated with `trec_eval` command in putty and appropriate map values are identified.

4. TUNING OF PARAMETERS

The tuning process involves the tweaking of parameters in order to improve the performance of the models.

4.1 Tuning B and K1 values in BM25 Model:


I have tried out tuning the parameters of BM25 model which has the default setting: $b = 0.75$ and $k1 = 1.2$. For the default setting, $MAP = 0.2451$. It is generally recommended that the values of b range from 0.5 to 0.8. I have used dismax as query parser and chose $b = 0.6$ and varied $k1$ from 0.1 to 1.5 as shown below in the table. At $k1 = 0.2$, A peak in the graph can be observed and the MAP value corresponding to this peak value is 0.6889.



The table for $k1$, b and map values is:

k1	b	MAP value
0.1	0.6	0.6862
0.2	0.6	0.6889
0.4	0.6	0.6866
0.7	0.6	0.6832
1	0.6	0.6766
1.5	0.6	0.6743

The screenshot of trec_eval program where MAP value of 0.6889 obtained at k1=0.2 and b=0.6 is

 k=0.2 b=0.6 - Notepad

File	Edit	Format	View	Help
iprec_at_recall_0.80	015	1.0000		
iprec_at_recall_0.90	015	0.0000		
iprec_at_recall_1.00	015	0.0000		
P_5	015	1.0000		
P_10	015	1.0000		
P_15	015	0.8667		
P_20	015	0.6500		
P_30	015	0.4333		
P_100	015	0.1300		
P_200	015	0.0650		
P_500	015	0.0260		
P_1000	015	0.0130		
runid	all	BM25		
num_q	all	15		
num_ret	all	280		
num_rel	all	225		
num_rel_ret	all	121		
map	all	0.6889		
gm_map	all	0.6168		
Rprec	all	0.6714		
bpref	all	0.6912		
recip_rank	all	1.0000		
iprec_at_recall_0.00	all	1.0000		
iprec_at_recall_0.10	all	0.9762		
iprec_at_recall_0.20	all	0.9333		
iprec_at_recall_0.30	all	0.9037		
iprec_at_recall_0.40	all	0.8698		
iprec_at_recall_0.50	all	0.7284		
iprec_at_recall_0.60	all	0.6447		
iprec_at_recall_0.70	all	0.5598		
iprec_at_recall_0.80	all	0.3778		
iprec_at_recall_0.90	all	0.3037		
iprec at recall 1.00	all	0.3037		

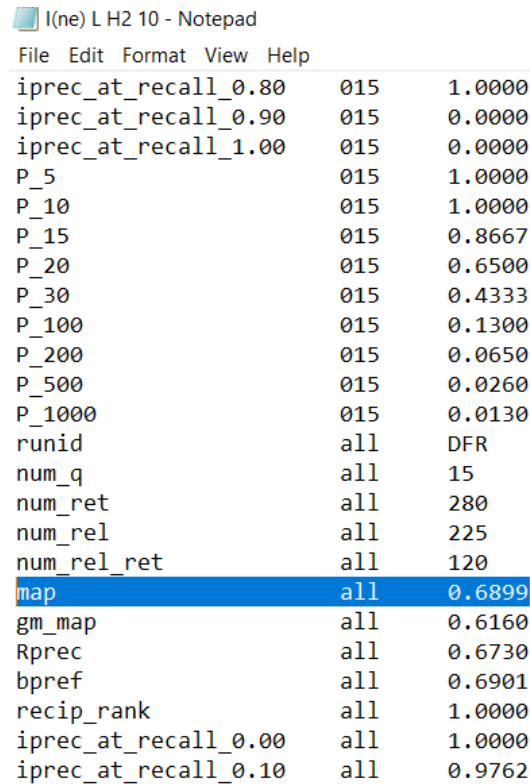
4.2 Tuning in DFR Model:

I have tried out various combinations of parameters such as normalization, aftereffect and basic model. When normalization – H2, Aftereffect – B and Basic model – G are used as default, the map value obtained is 0.2466. I have used defmax as the query parser and obtained a MAP value of 0.6899 when normalization-I(ne), Aftereffect-L and Basic model-H2 and constant c=10 are used.

The table for different values tried out is:

BasicModel	Aftereffect	Normalization	c	MAP value
G	B	H2		0.6752
I(ne)	L	H2	2	0.6852
I(ne)	L	H2	3	0.6756
G	B	H2	5	0.6771
I(n)	L	H2		0.6761
I(ne)	B	H2	500	0.6843
I(ne)	L	H2	10	0.6899
I(ne)	L	H2	50	0.6883

The screenshot of trec_eval program where MAP value of 0.6899 obtained for basic model=I(ne),aftereffect=L, Normalization=H2 and c=10 is



```

I(ne) L H2 10 - Notepad
File Edit Format View Help
iprec_at_recall_0.80    015    1.0000
iprec_at_recall_0.90    015    0.0000
iprec_at_recall_1.00    015    0.0000
P_5                    015    1.0000
P_10                   015    1.0000
P_15                   015    0.8667
P_20                   015    0.6500
P_30                   015    0.4333
P_100                  015    0.1300
P_200                  015    0.0650
P_500                  015    0.0260
P_1000                 015    0.0130
runid                  all    DFR
num_q                  all    15
num_ret                all    280
num_rel                all    225
num_rel_ret            all    120
map                    all    0.6899
gm_map                 all    0.6160
Rprec                  all    0.6730
bpref                  all    0.6901
recip_rank             all    1.0000
iprec_at_recall_0.00    all    1.0000
iprec_at_recall_0.10    all    0.9762

```

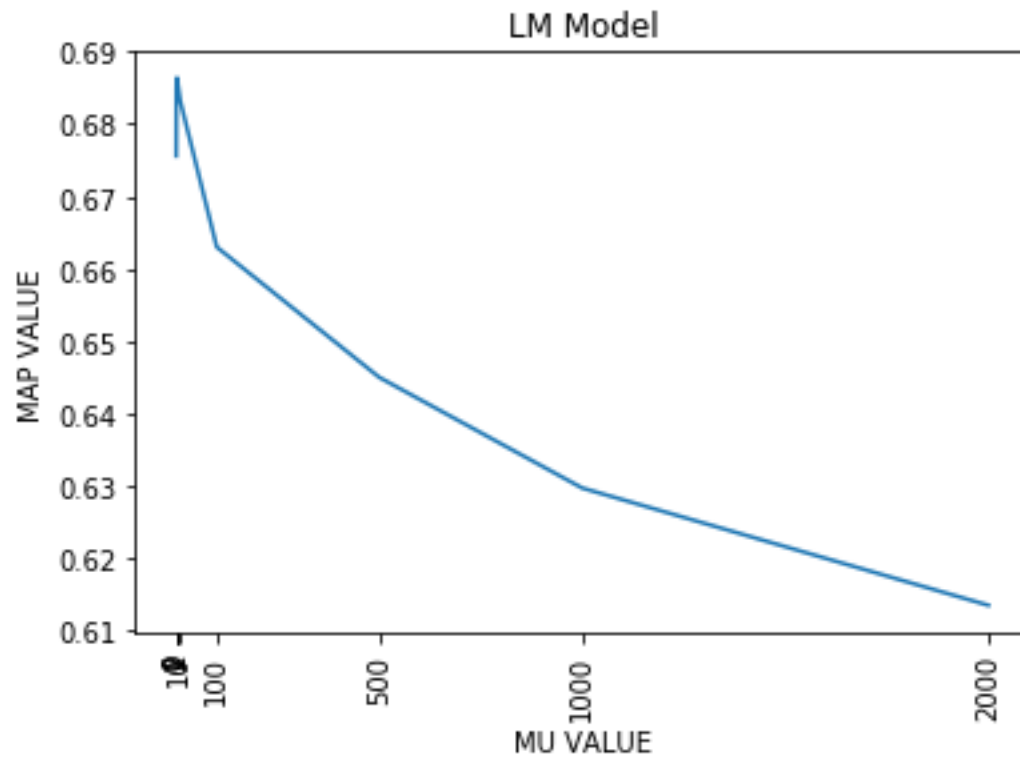
4.3 Tuning in LM Model:

When LM was implemented without any query parser I have obtained the MAP value as 0.2521. I have used defmax query parser in-order to improve the performance. At $\mu = 2$, A peak in the graph can be observed and the MAP value corresponding to this peak value is 0.6864.


The tabular form is as follows.

Constant value(mu)	MAP Value
0.1	0.6756
0.5	0.6850
1	0.6855
2	0.6864
10	0.6832
100	0.6631
500	0.6452
1000	0.6297
2000	0.6135

The graph plotted between mu value and MAP values is:



The screenshot of trec_eval program where MAP value of 0.6864 obtained mu=2 is:

 2 - Notepad

File	Edit	Format	View	Help
iprec_at_recall_0.80	015	1.0000		
iprec_at_recall_0.90	015	0.0000		
iprec_at_recall_1.00	015	0.0000		
P_5	015	1.0000		
P_10	015	1.0000		
P_15	015	0.8667		
P_20	015	0.6500		
P_30	015	0.4333		
P_100	015	0.1300		
P_200	015	0.0650		
P_500	015	0.0260		
P_1000	015	0.0130		
runid	all	LM		
num_q	all	15		
num_ret	all	280		
num_rel	all	225		
num_rel_ret	all	122		
map	all	0.6864		
gm_map	all	0.6185		
Rprec	all	0.6564		
bpref	all	0.6921		
recip_rank	all	1.0000		
iprec_at_recall_0.00	all	1.0000		

5. CONCLUSION

After enhancing search engine performance of the models, the following results have been obtained when defmax parser is used:

BM25 k1=0.2 b=0.6 MAP=**0.6889**

DFR basic-model=I(ne) aftereffect=L Normalization=H2 constant(c)= 10 MAP=**0.6899**

LM mu=2 MAP=**0.6864**