CS6501 Information Retrieval Homework

Tianye Song, ts7fx

4/9/17

1. Copy and paste your implementation of each ranking algorithm and evaluation function into your report, together with the corresponding final MAP/P@10/MRR/NDCG@10 performance you get from each ranking function. *Use the default parameter settings suggested*[*here*](http://www.cs.virginia.edu/~hw5x/Course/IR2017/_site/mps/2017/03/29/mp/#default) (30pts + 20pts)
   1. Ranking algorithm & Evaluation function implementations
   2. MAP, P@10, MRR, NDCG@10 performances

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | MAP | P@10 | MRR | NDCG@10 |
| Boolean Dot Product | 0.51 | 0.29 | 0.59 | 0.62 |
| TFIDF Dot Product | 0.56 | 0.31 | 0.67 | 0.68 |
| Pivoted Length Normalization | 0.61 | 0.34 | 0.71 | 0.72 |
| Okapi BM25 | 0.51 | 0.31 | 0.59 | 0.63 |
| Jelinek-Mercer | 0.029 | 0.0086 | 0.033 | 0.041 |
| Dirichlet Prior | 0.38 | 0.19 | 0.46 | 0.49 |

1. Please carefully tune the parameters in BM25 and Dirichlet prior smoothed Language Model. Report the best MAP you have achieved and corresponding parameter settings. (15pts)

For BM25 there are three parameters.. k1, k2 and b.

k1∈[1.2,2],k2∈(0,1000],b∈[0.75, 1.2]

try extreme values for k1, with k2 = 750 and b = 1

|  |  |  |  |
| --- | --- | --- | --- |
| K1 = | 1.2 | 1.5 | 2 |
| MAP | 0.539 | 0.509 | 0.482 |

Seems like k1 = 1.2 yield max MAP.

Now do k2’s values interfere MAP for k1= 1.2?

Try k2 = 1, 750, 1000, k1 = 1.2 and b = 1

|  |  |  |  |
| --- | --- | --- | --- |
| K2 = | 1 | 750 | 1000 |
| MAP | 0.539 | 0.539 | 0.539 |

K2 does not cause changes in MAP, so k2 values don’t matter for this case.

What about b? try b = 0.75, 1, 1.2, with k1 = 1.2 and k2 = 750

|  |  |  |  |
| --- | --- | --- | --- |
| b = | **0.75** | 1 | 1.2 |
| MAP | **0.595** | 0.539 | 0.478 |

Zoom into b = 0.75, with k1 = 1.2 and k2 = 750

|  |  |  |  |
| --- | --- | --- | --- |
| b = | **0.75** | 0.80 | 0.85 |
| MAP | **0.595** | 0.592 | 0.583 |

So parameter set that maximize MAP performance for ok BM25 model is k1 = 1.2 and b = 0.75. k2 can be anything within the range of (0,1000.

For Dirichlet Prior there is just one parameter, mu. The instruction mentions empirically the value for mu lies between 2000 to 3000.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 2000 | 2200 | 2400 | 2600 | 2800 | 3000 |
| MAP | 0.393 | 0.392 | 0.386 | 0.380 | 0.378 | 0.377 |

It is clear that since mu=2000, increasing mu leads to decreasing MAP.

What about mu less than 2000?

Try mu = 1800, mu = 1600, mu = 1400

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1400 | **1600** | 1800 |
| MAP | 0.395 | **0.398** | 0.394 |

So it seems like the peak MAP is somewhere around mu=1600.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1400 | 1500 | **1600** | 1700 | 1800 |
| MAP | 0.395 | 0.396 | **0.398** | 0.3963 | 0.394 |

Ok. Let’s keep zooming in.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1500 | 1550 | 1600 | **1650** | 1700 |
| MAP | 0.396 | 0.397 | 0.398 | **0.399** | 0.396 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1600 | 1620 | **1640** | 1650 | 1660 |
| MAP | 0.398 | 0.3985 | **0.3986** | 0.3985 | 0.3977 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1620 | 1630 | **1640** | 1650 |
| MAP | 0.3985 | 0.3985 | **0.3986** | 0.3985 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1635 | **1640** | **1645** |
| MAP | 0.3985 | **0.3986** | 0.3986 |

Thus, mu = 1640/1645 gives the best MAP for the DP model.

1. In edu.illinois.cs.index.SpecialAnalyzer.java, we defined a special document analyzer to process the document/query for retrieval purpose. Basically, we built up a pipeline with filters of LowerCaseFilter, LengthFilter, StopFilter, and PorterStemFilter. **Please disable some of the filters**, e.g., without stopword removal or stemming, and test the new analyzer with the **BM25** model (with your best parameters of step II). What is your conclusion about the effect of document analyzer on retrieval effectiveness? (15pts) ***Note: this analyzer has to be used in both indexing time and query time!***

Parameters: k1 = 1.2, k2 = 750, b = 0.75.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | MAP | P@10 | MRR | NDCG@10 |
| All filters active | 0.60 | 0.35 | 0.68 | 0.70 |
| No lowercase Filter | 0.60 | 0.35 | 0.68 | 0.70 |
| No LengthFilter | 0.60 | 0.35 | 0.68 | 0.70 |
| No stopFilter | 0.47 | 0.24 | 0.56 | 0.58 |
| No PorterStemFilter | 0.52 | 0.27 | 0.62 | 0.63 |
| Only lowercase and length filters | 0.40 | 0.19 | 0.46 | 0.50 |
| No filters at all | 0.41 | 0.19 | 0.47 | 0.51 |

For this given corpus, the stopFilter and PorterStemFilter are important. Taking any one of them out decrease the effectiveness of the BM25 retrieval model. Moreover, taking both of them out, the BM25 model performs even worse than if there are no filters at all.

1. With the default document analyzer, choose one or two queries, where TF-IDF dot-product model performed **significantly better** than Boolean dot-product model in average precision, and analysis what is the major reason for such improvement? Perform the same analysis for TF-IDF dot-product model v.s. BM25, and BM25 v.s. Dirichlet Prior smoothed Language Model and report your corresponding analysis (using your best parameters for BM25 and Dirichlet Prior smoothed Language Model). (10pts)

|  |  |  |  |
| --- | --- | --- | --- |
| Models compared | queries | Average precisions (tfidf vs bdp) |  |
| Tfidf vs bdp | measurement of plasma temperatures in arc discharge using shock wave techniques | 1.0 vs 0.25 | Some words (plasma, arc, shock) only appeared once |
| Tfidf vs bdp | variable ultra high frequency attenuators | 0.59 vs 0 | Some words(ultra, attenuators) only appeard once |
| Tfidf vs BM25 | measurement of plasma temperatures in arc discharge using shock wave techniques | 1.0 vs 0.5 | Same reason as above. Some words are rare. Thus tfidf does a better job in retrieving them |
| Tfidf vs BM25 | methods of calculating instantaneous power dissipation in reactive circuits | 0.83 vs 0.27 |  |
| BM25 vs DP | **characteristics of the single electrode discharge in the rare gases at low pressures** | 1.0 vs 0 |  |
| BM25 vs DP | active audio frequency filter with variable cut off slope 0.7634920634920634 | 0.76 vs 0 |  |

1. Pick one of the previously implemented scoring functions out of
   1. Okapi BM25
   2. Pivoted Length Normalization
   3. Langauge Model with Dirichlet Smoothing

to analyze under what circumstance the chosen scoring function will **mistakenly favor some less relevant document (*i.e.*, ranks a less relevant document than a more relevant one).**

After reading the paper [An Exploration of Axiomatic Approaches to Information Retrieval](http://www.eecis.udel.edu/~hfang/pubs/sigir05-axiom.pdf), how do you think you can fix the problem? Please relate your solution and corresponding implementation in the report. Also report the corresponding ranking performance of your revised ranking algorithm. (10pts)