Dublin City University
School of Computing
CA4009: Search Technologies
Laboratory Session 3

November 2016

Module Coordinator: Gareth Jones

Laboratory Tutors: Piyush Arora, Wei Li

1 Introduction

This lab extends the work on evaluation of information retrieval algorithms introduced in the previous lab by using the information retrieval test collections with the trec_eval program to examine and optimise the behaviour of ranked information retrieval using the BM25 probabilistic retrieval model.

2 Investigating BM25 Parameter Settings

2.1 Background

Your use of the BM25 algorithm for ranking the TREC collection documents using Lucene via the online search tool has, so far used a single set values for the parameters K and b. As explained in lectures, varying the values of K and b changes the weights of terms calculated using the BM25 algorithm. The changes in term weights produce consequential changes to the matching scores between the query and each document. The changes in matching scores can on some occasions change the ranks of the retrieved documents. Thus comparing the rank of retrieved particular documents for ranked lists created using BM25 with different K and/or b values, you will almost certainly find some of them at higher or lower positions in the ranked lists.

The objective in modifying the BM25 weights is to maximise expected retrieval effectiveness by seeking to increase the rank of relevant documents with a corresponding reduction in the rank of non-relevant documents[1]. In order to set the best values of K and b, the values of these parameters can be adjusted for a series of retrieval runs using an IR test collection. The retrieval behaviour for each run can be evaluated using the trec_eval program that you installed in the previous lab. The values of one or more of the experimental metrics (Precision, MAP, Recall, etc.) are typically compared for different parameter values, with the parameter values which produce the optimal metric values used as the operational settings.

2.2 Experimental Procedure

For this lab, you will carry out your own retrieval runs for BM25 ranking using Lucene with the information retrieval test collection you downloaded in the last lab consisting of: TREC topics sets, the Lucene index of

[1] Increase the rank here means put the document nearer the top of the list.

the TREC documents, and the relevance information for these topics contained in the qrel file. This is the same data you used in the previous lab, but used in a slightly different way.

To carry out this part of the laboratory:

- Using the interactive interface from the previous labs available at `http://136.206.115.117:8080/IRModelGenerator/`. select "Search" option and then select "Batch execute TREC queries" .
- This offers you 3 sets of different TREC queries. You should see a combo box where you can select which TREC query set you wish to use. Available options are TREC 6 (query ids 301-350), TREC 7 (query ids 351-400) and TREC 8 (query ids 401-450). Choose one set of topics for your first set of experiments. You can repeat the procedure for the other topic sets afterwards.

  These are the same topics that you downloaded before, split into sets of 50 topics. Although these topic files are stored in the server, it would be beneficial for you to manually inspect a few of them again to re-familairize yourself with the structure of the queries (each query contains three fields: title, description and narrative). The server makes use of only the title field for batch execution.

  After choosing your query set, selecting "Execute queries" passes each individual query to Lucene in turn which executes on retrieval on the server. Results are passed back to you in a single file which you can download.
- Select search with the "BM25" ranking option. You also need to set values of K and b. For your first run, you can use the same values as before $K = 1.2$ and $b = 0.75$. Subsequently, you need to carry out experiments changing these parameters and explore the effect on MAP (mean average precision) with parameter variations.
- To carry out a run, input the values of K and b in the respective text boxes and click "Execute Queries". The interface will display a busy spinner and you have to wait until the server finishes executing the 50 queries in the batch you selected. After this finishes, the server will return a link to the generated results file. You need to download this file by clicking on the link. Save the content in a local file on your computer. This file is in the same standard TREC results as the ranked lists results file that you examined in laboratory 2.
- You then need to evaluate the retrieval results in the downloaded file. To do this, run trec_eval with appropriate parameters, i.e. the *qrel* file (qrels.trec678.adhoc) and the retrieval results file that you saved locally. This TREC qrel file contains relevance information for all of the topics sets (TREC 6, TREC 7, TREC 8) . trec_eval ignores relevance information for topics in the qrel files which do not appear results file. The results for the run on the set of queries that you have run should then be displayed on the screen, in the same manner as they were for your previous use of trec_eval.

Once you have successfully carried out this procedure for a single pair of K and b parameter values you can experiment with some other parameter settings.

2.3 Optimisation of Parameter Values

To set up an information retrieval system for a particular task, developers typically carry out a series of runs with a systematic variation in the K and b values (or more generally a series of runs with a systematic variation of whatever parameters need to be set for the algorithm that they are using).

For the next part of the lab, you need to carry out and report an investigation of this sort.

- Your objective is to carry our a series of runs and to generate a line graph (in Excel or gnuplot or your alternative favourite package for generating a line graph) which shows the MAP values on the y-axis and K values on the x-axis (vary K from 0.2 to 2 in steps of 0.2). Each line in this graph should represent a set of results with a fixed b value (b = 0.1 to 0.9 in steps of 0.1). Based on the results in the graph you should be able to recommend the best values of K and b to use for BM25 with your test collection.
- This information can also be plotted on a 3D graph with K and b on the x-axis and the y-axis and MAP on the z-axis. This produces a single plane which enables you to see the interrelationship between K and b. If you have time, you can try to create a plot of this type from which you might find it easier to select your K and b values.

You should include your plots in a written report, including comments on what you observe in the lines on the graphs, your recommended values for K and b, with a short explanation of why you chose these values.

In your analysis, you should consider what it means in terms of the term weights, and the impact of these terms weights on the retrieval results when you change the values of K and b. To do this, you will probably find it helpful to refer to the lecture notes on term weighting in the Section 3: Text Retrieval and the additional note on BM25 on the loop page, and consider the role of K and b in determining the term weights and the purpose of each of these parameters.

Your report should be submitted to the laboratory 3 turnitin link.

You can carry out all these operations manually or automate the process. The next section gives details of how to automate the process. If you are confident with these methods, you could take this approach straight away. Otherwise complete these task manually, and investigate the automated method afterwards if you wish. Additional credit will be given for successfully completing the automated method.

When you have created the plots for all three topic sets, compare them. Are they the same? Do they suggest different optimal K and b values? If they do, how would you choose the K and b values for an online search system based on BM25?

2.4 Optional: Advanced Section: Automating the Procedure

While you can vary the parameters for this task using the interface and evaluate the runs manually, an easier and faster way to achieve this is to use a script to automate this process.

The service (called automatically by the interface) which does the computation for you is `http://136.206.115.117:8080/IRModelGenerator/TrecBatchQueryExecuterServlet`.

To operate a script to to call this service, you need to call it with appropriate HTTP GET arguments specifying the query set on which to execute queries, the similarity function to use and the parameter values for K and b. For example,

`http://136.206.115.117:8080/IRModelGenerator/TrecBatchQueryExecuterServlet?treccode=6&simf=BM25&k=1&b=0.9`

executes TREC 6 queries using BM25 as the retrieval model with K and b set to 1 and 0.9 respectively.

To use this more automated approach you need to write a small program or script to generate these URL values to hit the online retrieval service. To call the service automatically, one option is to use a shell script and make repeated calls to *wget*. The next step is to hit the URL which this page returns (which contains the results file). You can then redirect the results file content to a file from within your script and make a call to trec_eval. Each MAP value obtained with a particular parameter setting can then be saved to a comma

separated file (grep -w “MAP” on the trec_eval output would give you the MAP value), which can then be opened in Excel (or OpenOffice) to generate the line plot for parameter variation effect on MAP. You can also make use of gnuplot (if you are familiar with it).

If you complete this automated approach. Please submit details of how you did this, and submit your script via turnitin.