On the Relationship Between Effectiveness and Accessibility

Leif Azzopardi
Department of Computing Science
University of Glasgow
United Kingdom
leif@dcs.gla.ac.uk

Richard Bache
Department of Computing Science
University of Glasgow
United Kingdom
bache@dcs.gla.ac.uk

ABSTRACT

Typically the evaluation of Information Retrieval (IR) systems is focused upon two main system attributes: efficiency and effectiveness. However, it has been argued that it is also important to consider accessibility, i.e. the extent to which the IR system makes information easily accessible. But, it is unclear how accessibility relates to typical IR evaluation, and specifically whether there is a trade-off between accessibility and effectiveness. In this poster, we empirically explore the relationship between effectiveness and accessibility to determine whether the two objectives i.e. maximizing effectiveness and maximizing accessibility, are compatible, or not. To this aim, we empirically examine this relationship using two popular IR models and explore the trade-off between access and performance as these models are tuned.

Categories and Subject Descriptors: H.3.3 Information Storage and Retrieval - *Retrieval Models*

General Terms: Theory, Experimentation

Keywords: Information Retrieval, Accessibility, Findability, Retrievability, Evaluation

1. INTRODUCTION

Historically, there have been two main ways to evaluate an Information Retrieval (IR) system: efficiency and effectiveness [7]. A complementary and so-called higher order evaluation has been recently proposed based on accessibility [1]. Instead of assessing how well the system performs in terms of speed or performance, access-based measures provide an indication of how easily documents within the collection can be retrieved using a particular retrieval system [1]. Evaluations based on accessibility have been performed in a number of different contexts (see [2, 6, 3, 4]), but there has been little work examining the relationship between access-based measures and effectiveness measures.

2. MEASURING ACCESSIBILITY

The accessibility of information in a collection given a system has been considered from two points of view, the system side i.e. *retrievability* [2] and the user side *findability* [6]. Retrievability measures provide an indication of how easily a

document could be retrieved using a given IR system, while findability measures provide an indication of how easily a document can be found by a user with the IR system. Here we consider the accessibility measures based on retrievability (see [2] for more details and [3, 4] for examples of its usage in practice.)

Retrievability: The retrievability $r(d)$ of a document d with respect to an IR system, is the ease with which it can be retrieved given all possible queries Q[1]. Formally,

$$r(d) \propto \sum_{q \in Q} f(k_{dq}, c) \qquad (1)$$

where q is a query in Q, k_{dq} is the rank at which d is retrieved given q, and $f(k_{dq}, c)$ is a function which denotes how accessible the document d is for the query q given the rank cutoff c. A simple measure of retrievability employs an access function $f(k_{dq}, c)$, such that if d is retrieved in the top c documents given q, then $f(k_{dq}, c) = 1$ else $f(k_{dq}, c) = 0$. This is referred to as a cumulative-based retrievability measure [2]; and it provides an intuitive value for each document, i.e. it is the number of times that the document is retrieved in the top c documents. To provide a single measure of access given the retrievability scores for all documents, the Gini measure [5] was proposed in [2]. Intuitively, if all documents were equally accessible according to $r(d)$, then the Gini would be zero (equality), while if all but one document had $r(d) = 0$ then the Gini would be one (total inequality). Usually, most documents have some level of retrievability and the Gini measures is somewhere between zero and one. Essentially, the Gini coefficient provides an indication of the level of inequality between documents given how easily they can be retrieved with a particular retrieval system.

It is important to note that retrievability can be estimated without recourse to relevance judgments, making it an attractive alternative for automatic evaluation. That is, if there is a positive correlation between accessibility and effectiveness based measures.

Relating Retrievability and Effectiveness: Given the definition of retrievability then a purely random IR system would provide equal access to all documents (i.e. Gini=0); however, this would also result in very poor effectiveness. While, if a (hypothetical) IR system retrieved the set of known relevant documents, and only these documents, regardless of the query, then there would be a very high inequality among documents and Gini would be close to one.

---

[1]Since we cannot know the set of all the possible queries, it is usually approximated using a large (order of 10^5) set of automatically generated queries[2, 3, 4].

Neither extreme is desirable, but to what extent do we need to trade-off retrievability for effectiveness. In this poster, we examine the relationship between the Gini measure (i.e. the summarized retrievability measure) and precision, by examining the change in each measure as the parameter values of different retrieval models are varied.

Experimental Setup Two TREC collections were used: Associated Press (AP) 1988-1989 and Wall Street Journal (WSJ) 1987-1992, both with TREC query sets 1, 2 and 3. Two popular IR models were selected: Multinomial Language Modelling with Bayes Smoothing and Okapi BM25. For Language modelling, the smoothing parameter β was varied from 10^{-4} to 10^{5} in multiplicative steps of 10. For BM25, the b parameter, which adjusts length normalisation, was varied from 0.1 to 1.0 in steps of 0.1. Effectiveness was measured using both precision at 10 (P@10) and mean average precision (MAP). To estimate retrievability, the same methodology employed in [2] was applied, where we use 100,000 two-word queries derived from the most frequent collocations found in each corpus to estimate the retrievability values. Retrievability was measured using the cumulative measure (described above) where $c = 10$ and $c = 100$. The degree of equality was measured using the Gini coefficient denoted as Gini@10 and Gini@100, respectively.

Results In Figure 1, plots of the different measures are shown for each model (top: Language Model, bottom: BM25) and each collection (left: AP, right: WSJ) across the parameter values. From these plots, the first point of interest is that Gini varies considerably across the parameter ranges for both models, where minimizing the Gini coefficient translates into providing more access to documents in the collection (this is around $\beta = 10 - 100$ for the Language Model and $b = 0.6 - 0.8$ for BM25. Of note, is that the suggested/default value for b is usually 0.75, which is well within this range.). While this does not directly correspond to where performance is maximized, the difference in performance is quite small; in the range of (0.01-0.03 for both P@10 and MAP). While, these differences in performance were significantly difference ($p < 0.05$ using Student's T-test) for all but BM25 on WSJ, there does appear to be a positive correlation between the two measures, and this opens up the possibility of using access based measures to tune retrieval systems. These findings suggest that a systematic relationship appears to exist between the gini measurements (representing Accessibility) and the precision measurements (representing Effectiveness).

3. DISCUSSION AND FURTHER WORK

This preliminary analysis of the relationship between accessibility measures (specifically retrievability measures) and effectiveness measures shows that the two goals of maximizing access and maximizing performance are quite compatible. In fact, reasonably good retrieval performance is still obtained by selecting parameters that maximize access (i.e. when there is the least inequality between documents according to Gini given the $r(d)$ values). This motivates the hypothesis that retrieval models/systems can be effectively tuned using access based measures. If this holds, then it suggests that when relevance information is not available a sensible approach to configuring a system is to ensure that users can access all documents as easily as possible. However, further research is needed to test this hypothesize more

conclusively and to explore this interesting and complex relationship in detail. In future work, we shall examine different models, and the influence of the different parameters on access and performance.
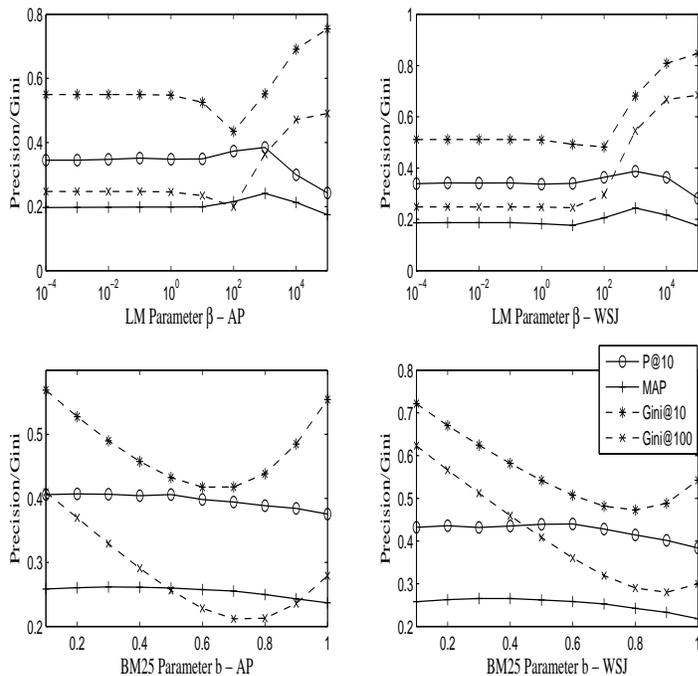
Figure 1: Plots of Precision and Gini measures across parameters for the LM and BM25 models.

4. REFERENCES

[1] L. Azzopardi and V. Vinay. Accessibility in IR. In *Proceedings of 30th ECIR 2008*, 482–489, 2008.

[2] L. Azzopardi and V. Vinay. Retrievability: An evaluation measure for higher order information access tasks. In *Proceedings of the 17th ACM CIKM 2008*, 561–570, 2008.

[3] S. Bashir and A. Rauber. Improving retrievability of patents with cluster-based pseudo-relevance feedback documents selection. In *Proceedings of the 18th ACM CIKM 2009*, 1863–1866, 2009.

[4] S. Bashir and A. Rauber. Improving retrievability of patents in prior-art search. In *Proceedings of 32nd ECIR 2010*, 457–470, 2010.

[5] J. Gastwirth. The estimation of the lorenz curve and gini index. *The Review of Economics and Statistics*, 54:306–316, 1972.

[6] H. Ma, R. Chandrasekar, C. Quirk, and A. Gupta. Improving search engines using human computation games. In *Proceedings of the 18th ACM CIKM 2009*, 275–284, 2009.

[7] C. J. van Rijsbergen. *Information Retrieval*. Butterworths, London, 1979.