Re-Ranking Search Results Using Document-Passage Graphs

Michael Bendersky
Center for Intelligent Information Retrieval
Department of Computer Science
University of Massachusetts
Amherst, MA 01003
bemike@cs.umass.edu

Oren Kurland
Faculty of Industrial Engineering and Management
Technion – Israel institute of technology
Technion City, Haifa 32000, Israel
kurland@ie.technion.ac.il

ABSTRACT

We present a novel *passage*-based approach to re-ranking documents in an initially retrieved list so as to improve precision at top ranks. While most work on passage-based document retrieval ranks a document based on the query similarity of its constituent passages, our approach leverages information about the *centrality* of the document passages with respect to the initial document list. Passage centrality is induced over a bipartite document-passage graph, wherein edge weights represent document-passage similarities. Empirical evaluation shows that our approach yields effective re-ranking performance. Furthermore, the performance is superior to that of previously proposed passage-based document ranking methods.

Categories and Subject Descriptors: H.3.3 [Information Search and Retrieval]: Retrieval Models

General Terms: Algorithms, Experimentation

Keywords: passage-based document retrieval, passage language models, centrality, passage-document graphs

1. INTRODUCTION

To improve precision at the top ranks of results returned in response to a query, researchers suggested to automatically re-rank the documents in an initially retrieved list [11, 7, 2, 4, 5]. Information induced from inter-document similarities is often used in these approaches. Specifically, document *centrality*, as measured by textual similarity to (many) other (central) documents in the initial list (or their clusters), was effectively utilized for re-ranking [4, 5].

An issue not accounted for in the re-ranking approaches just mentioned is that long and/or heterogeneous relevant documents may contain very few parts (*passages*) that pertain to the query. Indeed, methods for ranking all documents in a corpus, which utilize passage-query similarity information, were designed to address this issue [9, 1, 6].

We propose a novel approach to the re-ranking task that leverages insights from the two lines of work just described. We (re-)rank the documents in the initial list by utilizing information about their constituent passages' centrality with respect to the document list. Passage centrality is defined (using graph-based methods) in terms of similarity to (central) documents in the list — analogously to a cluster-centrality definition in past work [5]. Our hypothesis is that passages similar to many documents in the list contain information that pertains to the query due to the virtue by which the list was created. (That is, in response to the query.)

We show that our approach is effective in re-ranking the initial document list, and that it posts better re-ranking performance than that of previous document ranking methods that utilize passage-query similarity information.

SIGIR'08, July 20–24, 2008, Singapore.
ACM 978-1-60558-164-4/08/07.

2. GRAPH-BASED RE-RANKING

In what follows we assume that the following have been fixed: a query q, a set $\mathcal{D}_{\text{init}}$ of the documents most highly ranked by some initial search performed in response to q, and a set $PG(\mathcal{D}_{\text{init}})$ of all passages of documents from $\mathcal{D}_{\text{init}}$. To construct a document-passage graph, we adopt an approach for creating cluster-document graphs [5].

We define a complete weighted directed graph $G = (V, wt)$ over $V = \mathcal{D}_{\text{init}} \cup PG(\mathcal{D}_{\text{init}})$, where wt is an edge-weight function that is based on a similarity measure $sim(x, y)$:

$$wt(u,v) \stackrel{def}{=} \begin{cases} sim(u,v) & \text{if } u \in \mathcal{D}_{\text{init}} \text{ and} \\ & \quad v \in Nbhd(u \mid \delta, PG(\mathcal{D}_{\text{init}})), \\ 0 & \text{otherwise;} \end{cases}$$

$Nbhd(u \mid \delta, PG(\mathcal{D}_{\text{init}}))$ is the set of δ passages g in $PG(\mathcal{D}_{\text{init}})$ that yield the highest $sim(u, g)$. We set $sim(x, y) \stackrel{def}{=} \exp\left(-D\left(p_x^{[0]}(\cdot) \,\middle|\middle|\, p_y^{[\mu]}(\cdot)\right)\right)$, where D is the KL divergence and $p_z^{[\mu]}(\cdot)$ is the unigram Dirichlet-smoothed language model induced from z (μ is the smoothing parameter) [12].

Thus, G is essentially a one-way bipartite graph in which an edge with a non-zero weight connects document d in $\mathcal{D}_{\text{init}}$ with the δ passages in $PG(\mathcal{D}_{\text{init}})$ that are most similar to d.

To determine the centrality of passage g ($Cent(g)$), we either compute its **influx**: $\sum_{d \in \mathcal{D}_{\text{init}}} wt(d, g)$, or use its **authority** score as induced by the HITS (hubs and authorities) algorithm [3], when considering the weights of edges. (Note that documents are hubs and passages are authorities in G.)

We then rank document $d \in \mathcal{D}_{\text{init}}$ by using the centrality of its most central passage as a bias on d's query-similarity[1]:

$$Score(d) \stackrel{def}{=} sim(q,d) \max_{g_i \in d} Cent(g_i), \qquad (1)$$

to address cases wherein d is long (heterogeneous) and contains very few passages that pertain to q. (We write $g_i \in d$ to indicate that passage g_i is part of document d.)

[1] Using only passage centrality yields degraded performance.

corpus	queries	disks
AP	51-64, 66-150	1-3
TREC8	401-450	4-5
WSJ	151-200	1-2

	AP			TREC8			WSJ		
	p@5	p@10	MRR	p@5	p@10	MRR	p@5	p@10	MRR
DocBase	45.7	43.2	59.6	50.0	45.6	69.1	53.6	48.4	74.8
PsgBase	46.1	41.7	60.2	44.8^{d}	43.0	63.9	48.8^{d}	44.6	67.7
InterPsgDoc	46.1	41.7	60.2	50.4^{p}	46.0^{p}	70.1	54.0^{p}	48.8^{p}	$\mathbf{78.2}^{p}$
MultPsgDoc	45.3	43.4	62.1	49.6	46.4^{p}	67.1	52.8^{p}	47.8^{p}	75.7_{i}
influx	$\mathbf{50.7}^{d}$	46.7^{dp}_{im}	60.4	55.2^{dp}_{im}	47.8^{p}	$\mathbf{72.5}_{m}$	$\mathbf{55.6}^{p}$	$\mathbf{50.8}^{p}$	70.6
authority	50.3	$\mathbf{47.3}^{dp}_{i}$	**60.4**	$\mathbf{55.6}^{p}$	**48.2**	68.2	53.2	49.2	67.7_{i}

Figure 1: Comparison of passage-centrality-based re-ranking methods (influx and authority) with the initial ranking of the list (DocBase) and re-ranking methods that utilize query-passage similarity information (PsgBase, InterPsgDoc and MultPsgDoc). Boldface: best result in a column; statistically significant differences with DocBase, PsgBase, InterPsgDoc and MultPsgDoc are marked with d, p, i, and m, respectively.

3. EVALUATION

We use the TREC corpora from Figure 1 for evaluation. The data is tokenized and Porter-stemmed using the Lemur toolkit (www.lemurproject.org), which is also used for language model induction. Topics' titles are used for queries.

The list upon which re-ranking is performed, $\mathcal{D}_{\text{init}}$, is set to the 50 documents in the corpus that are assigned the highest *initial ranking* score $sim(q,d)$. (For the purpose of estimating $sim(q,d)$, the document language model smoothing parameter, μ, is set here and after to a value optimizing map@1000, as in some previous work on re-ranking [4, 5].) $PG(\mathcal{D}_{\text{init}})$ is the set of half overlapping window passages [1, 6] (of 150 terms) of the documents in $\mathcal{D}_{\text{init}}$.

For reference comparisons to our centrality-based methods, we use **DocBase** — the initial ranking from which $\mathcal{D}_{\text{init}}$ is derived, and two commonly used passage-based document ranking approaches [9, 1, 6]: **PsgBase**, which scores $d \in \mathcal{D}_{\text{init}}$ by $\max_{g_i \in d} sim(q, g_i)$, and **InterPsgDoc**, which scores $d \in \mathcal{D}_{\text{init}}$ by $\lambda sim(q,d) + (1-\lambda) \max_{g_i \in d} sim(q, g_i)$; λ is a free interpolation parameter. (Note that PsgBase is a specific instance of InterPsgDoc with $\lambda = 0$.) We also explore a variant of Equation 1, denoted **MultPsgDoc**, which uses passage-query similarity information instead of passage-centrality information for scoring $d \in \mathcal{D}_{\text{init}}$: $sim(q,d) \max_{g_i \in d} sim(q, g_i)$.

We use precision at the top 5 and 10 documents (p@5, p@10) and the mean reciprocal rank of the first relevant document (MRR) to evaluate the effectiveness of the re-ranking methods in improving precision at top ranks. The free parameters of the re-ranking methods are set to values optimizing p@5: δ, the graph "out-degree", is chosen from $\{9, 19, \ldots, 99\}$, and λ (in the InterPsgDoc algorithm) is selected from $\{0, 0.1, \ldots, 1\}$; μ is set to 2000 [12]. Statistically significant performance differences are determined using the Wilcoxon two-tailed test at a confidence level of 95%.

We can see in Figure 1 that our passage-centrality-based methods are effective in re-ranking the initial list. (Compare the performance of influx and authority with that of DocBase — the initial document ranking.) Furthermore, our centrality-based methods are superior in most relevant comparisons (corpus $\times$ evaluation measure) to the reference comparisons that utilize passage-query similarity information (PsgBase, InterPsgDoc and MultPsgDoc); in many cases the performance differences are also statistically significant. In comparing the centrality-induction approaches (influx and authority), we see that none dominates the other; the performance differences between the two are not statistically significant.

4. CONCLUSION AND FUTURE WORK

Utilizing passage-centrality information is effective for re-ranking documents in a list retrieved in response to a query. Furthermore, the performance is superior to that of previous document ranking methods that utilize passage-query similarity information. We plan to explore centrality-induction methods over passage-solely graphs, which were used in work on sentence retrieval for question answering [8] and query-by-example retrieval [10]. We will also adopt document-centrality induction techniques [4, 5] for inducing passage-centrality, as some of these post re-ranking performance that is superior to that of our approach.

Acknowledgments We thank the reviewers for their comments. This paper is based upon work done in part while the first author was at the Technion and upon work supported in part by the Center for Intelligent Information Retrieval, the Jewish Communities of Germany Research Fund, and by a gift from Google. Any opinions, findings and conclusions or recommendations expressed in this material are the authors' and do not necessarily reflect those of the sponsors.

5. REFERENCES

[1] J. P. Callan. Passage-level evidence in document retrieval. In *Proceedings of SIGIR*, pages 302–310, 1994.

[2] F. Diaz. Regularizing ad hoc retrieval scores. In *Proceedings of the Fourteenth International Conference on Information and Knowledge Managment (CIKM)*, pages 672–679, 2005.

[3] J. Kleinberg. Authoritative sources in a hyperlinked environment. Technical Report Research Report RJ 10076, IBM, May 1997.

[4] O. Kurland and L. Lee. PageRank without hyperlinks: Structural re-ranking using links induced by language models. In *Proceedings of SIGIR*, pages 306–313, 2005.

[5] O. Kurland and L. Lee. Respect my authority! HITS without hyperlinks utilizing cluster-based language models. In *Proceedings of SIGIR*, pages 83–90, 2006.

[6] X. Liu and W. B. Croft. Passage retrieval based on language models. In *Proceedings of (CIKM)*, pages 375–382, 2002.

[7] X. Liu and W. B. Croft. Cluster-based retrieval using language models. In *Proceedings of SIGIR*, pages 186–193, 2004.

[8] J. Otterbacher, G. Erkan, and D. R. Radev. Using random walks for question-focused sentence retrieval. In *Proceedings of HLT/EMNLP*, pages 915–922, 2005.

[9] G. Salton, J. Allan, and C. Buckley. Approaches to passage retrieval in full text information systems. In *Proceedings of SIGIR*, pages 49–58, 1993.

[10] X. Wan, J. Yang, and J. Xiao. Towards a unified approach to document similarity search using manifold-ranking of blocks. *Information Processing and Management*, 44(3):1032–1048, 2008.

[11] P. Willett. Query specific automatic document classification. *International Forum on Information and Documentation*, 10(2):28–32, 1985.

[12] C. Zhai and J. D. Lafferty. A study of smoothing methods for language models applied to ad hoc information retrieval. In *Proceedings of SIGIR*, pages 334–342, 2001.