Australian Journal of Basic and Applied Sciences, 9(10) Special 2015, Pages: 125-130

ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

Australian Journal of Basic and Applied Sciences AENSI Publisher AJBAS

Using Base Line Methods Ranking on Data with Sink Points

[1]B.Yuvaraj and [2]Dr. C. Sureshganadhas

[1]*Research Scholar, Department of Computer Science and Engineering, Manonmaiam Sundaranar University, Tamil Nadu, India*
[2]*Professor & Head, Department of Computer Science and Engineering, Vivekanandhan College of Engineering for Women, Elayampalayam, Thiruchencode, Tamil Nadu*

ARTICLE INFO

Article history:
Received 28 January 2015
Accepted 25 February 2015
Available online 6 March 2015

Keywords:
Diversity in Ranking, Manifold Ranking with Sink Points, Update Summarization, Query Recommendation.

ABSTRACT

Ranking is a most important problem in various application domains, such as information retrieval, natural language processing, computational ecology, and social sciences. Many ranking approaches have been projected to rank objects according to their degrees of relevance or importance. Beyond these two goals, assortment has also been recognized as a critical criterion in ranking. Top ranked results are probable to convey as little redundant information as possible, and cover as many aspects as possible. However, present ranking approaches either take no account of diversity, or handle it individually with some heuristics. In this paper, we introduce a novel approach, Manifold Ranking with Sink Points (MRSP), to address assortment as well as relevance and importance in ranking. Specifically, our approach uses a manifold ranking process over the data manifold, which can logically find the most relevant and important data objects. Meanwhile, by spinning ranked objects into sink points on data manifold, we can efficiently prevent redundant objects from receiving a high rank. MRSP not only shows a nice convergence property, but also has an exciting and satisfying optimization clarification. We applied MRSP on two application tasks, update summarization and query suggestion, where diversity is of great concern in ranking. Experimental results on both tasks present a strong empirical performance of MRSP as compared to existing ranking approaches.

To Cite This Article: B.Yuvaraj and Dr. C. Sureshganadhas., Using Base Line Methods Ranking on Data with Sink Points. ***Aust. J. Basic & Appl. Sci.,*** *9(10): 125-130, 2015*

INTRODUCTION

RANKING has profuse applications in information retrieval (IR), data mining, and natural language processing. In various real circumstances, the ranking problem is defined as follows. Given a group of data objects, a ranking model (function) sorts the objects in the group conferring to their degrees of relevance, importance, or preferences (Lan, Y., 2009). For example, in IR, the "group" resembles to a query, and "objects" correspond to documents associated with the query. However, a mass of relevant objects may contain exceedingly redundant, even replicated information, which is disagreeable for users. Furthermore, the user's needs might be multi-faceted or uncertain. The severance in top ranked results will reduce the unintended to satisfy diverse users. For example, given a query "dirigible", if the top classified search results were all similar articles about the "Dirigible iPod speaker", it would be a waste of the output space and largely degrade users' search experience even though the results are all highly related to the query. Obviously, such top ranked results would not satisfy the users who want to know about the rigid airship "Zeppelin" or the rock band "Zeppelin". Thus, it is important to reduce redundancy in these top search results.

Therefore, beyond relevance and importance, diversity has also been recognized as a crucial criterion in ranking. Top ranked results are expected to carry as little redundant information as possible, and cover as many features as possible. In this way, we are able to reduce the risk that the information need of the user will not be satisfied. Many real application tasks demand assortment in ranking. For example, in query commendation, the suggested queries should capture different query intents of diverse users. In text summarization, candidate sentences of a summary are expected to be less redundant and cover different aspects of information delivered by the document. In e-commerce, a list of relevant but distinctive products is useful for users to browse and make a purchase.

(**MRSP**), to address assortment as well as relevance and importance in a incorporated way.

Corresponding Author: B.Yuvaraj, Research Scholar, Department of Computer Science and Engineering, Manonmaiam Sundaranar University, Tamil Nadu, India
E-mail: byuvarajb@gmail.com

Specifically, our approach uses a diverse ranking process (Zhu, X., *et al*., 2007; Zhu, X., 2010) over data manifold, which can help find the most relevant and important data objects. In the interim, we introduce into the manifold sink points, which are objects whose ranking scores are fixed at the least score (zero in our case).

2. Related Work:

2.1. Ranking on Data Manifolds:

The manifold ranking algorithm is anticipated based on the following two key expectations: (1) nearby data are likely to have close ranking scores; and (2) Data on the same structure are likely to have close ranking scores. An intuitive explanation of the ranking algorithm is described as follows. A partisan network is constructed first, where nodes represent all the data and query arguments, and an edge is put among two nodes if they are "close". Query nodes are then commenced with a positive ranking score, while the nodes to be ranked are allocated with a zero initial score. All the nodes then propagate their ranking scores to their neighbors via the weighted network. The dissemination process is repetitive until a global stable state is achieved, and all the nodes except the queries are ranked conferring to their final scores. The detailed ranking algorithm can be found in (Zhu, X., 2010).

2.2 Diversity in Ranking:

Beyond significance and importance, diversity has also been recognized as a crucial criterion in ranking recently (Radev, D.R., 2000; Zhang, J., 2008; Zhou, D., 2004;Li, W., 2008;Wen, J.R., 2001). Among the existing work, a well-known approach on introducing diversity in ranking is MMR (Carbonell, J. and J. Goldstein, 1998), which constructs a ranking metric combining the criteria of consequence and diversity, but leaving importance imprudent. Grasshopper addresses the problem by applying an absorbing random walk, but it has to leverage two different metrics to generate a diverse ranking list. Another work is Div Rank (Mei, Q., 2010), which uses a vertex-reinforced indiscriminate walk to introduce the rich-get-richer mechanism for diversity. However, topic relevance is not taken into account in this model. To the best of our knowledge, the challenge of addressing relevance, significance and diversity simultaneously in a unified way is still far from being well-resolved.

Manifold Ranking With Sink:
Points:

3.1 Main Idea:

In this paper, we propose a novel approach MRSP to address diversity as well as relevance and importance in ranking in a unified way. Specifically, MRSP assumes all the data and query objects are points sampled from a low-dimensional manifold and leverages a manifold ranking process (Zhu, X., *et al*., 2007; Zhu, X., 2010) to address relevance and importance.

Our overall algorithm follows an iterative structure. At each iteration, we use manifold ranking to find one or more most relevant points. Then, we turn the ranked points into sink points, update scores, and repeat. By turning ranked objects into sink points on data manifold, we can effectively prevent redundant objects from receiving a high rank. Note here that the key idea of MRSP is similar to absorbing random walk. However, absorbing random walk does not have the manifold assumption and it uses two different measures, *stationary distribution* and *expected number of visits before absorption*, to select the top ranked object and the remaining objects. This is largely different from MRSP where all the objects are ranked and selected using one consistent measure (i.e., the ranking score) based on the intrinsic manifold structure.

3.2 An Illustrative example:

We illustrate the proposed MRSP algorithm based on an example to show how it works. We created a dataset with 100 points as shown in Fig. 1(a). There are roughly three groups with

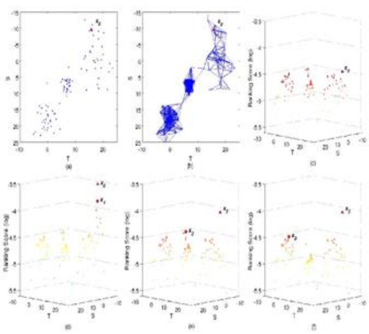

Fig. 1: (a) A data set. (b) The connected weighted network. (c) Ranking Score distribution when no prior knowledge on any point. (d) Ranking score distribution given Topic x_0. (e) Ranking score distribution given Topic x_0 and sink point x_1 . (f) Ranking score distribution given Topic x_0 and sink points x_1 and x_2.

3.3 The Algorithm and Its Convergence:

We now describe our MRSP algorithm in detail. Let $\chi = \chi_q \cup \chi_s \cup \chi_r \subset R^m$ denote a set of data points over the manifold, where $\chi_q = \{x_1, \ldots, x_q\}$ denotes a set of query points, $\chi_s = \{x_1, \ldots, x_s\}$ denotes a set of sink points, and $\chi_r = \{x_1, \ldots, x_r\}$ denotes the set of points to be ranked, called *free points*. Let f: $\chi \rightarrow$ R denote a ranking function which assigns a ranking score f_i to each point x_i . We can view f as a vector f $= [f_1 \ldots f_N]^T$, where N = q + s + r. We also define a vector $y = [y_1, \ldots, y_N]^T$, in which $y_i = 1$ if x_i is a query, and $y_i = 0$ otherwise. Suppose only top-K ranked data points are needed to be diversified, the MRSP algorithm works as follows:

1. Initialize the set of sink points _s as empty.
2. Form the affinity matrix W for the data manifold, where Wij = sim(xi, xj) if there is an edge linking xi and xj . Note that sim(xi, xj) is the similarity between objects xi and xj .
3. Symmetrically normalize W as $S = D^{-1/2}W^{D-1/2}$ in which D is a diagonal matrix with its (i, i)-element equal to the sum of the i-th row of W.
4. Repeat the following steps if |_s| < K:

(a). Iterate $f(t + 1) = \alpha SI_f\ f(t) + (1 - \alpha\_)y$ until convergence, where $0 \leq \_ < 1$, and If is an indicator matrix which is a diagonal matrix with its (i, i)-element equal to 0 if $x_i \in x_s$ and 1 otherwise.

(b). Let f_i* denote the limit of the sequence $\{f_i(t)\}$.Rank points xi $\in$ r according to their ranking scores f_i* (largest ranked first).

(c) Pick the top ranked point x_m. Turn x_m into a new sink point by moving it from X_r to X_s.

5. Return the sink points in the order that they were selected into X_s from X_r.

3.4 The Refined MRSP Algorithm:

Based on above analysis, the refined algorithm of MRSP is described in Fig. 2. Note that in step 5, matrix is initially organized by grouping sink points into Ω_{11} and others into Ω_{22} If the set of sink points is empty, we have $\Omega_{22} = \Omega$ which means the refined algorithm degenerates into the traditional manifold ranking algorithm. At each iteration, we mark the top-ranked object as a new sink point and move it from the group of free points to the group of sink points by reorganizing matrix Ω. Then the object to be selected next will deliver different information from that of already selected. With small number of query points in most real scenarios, the computation in step 6 can be very economical. In this way, our refined MRSP algorithm is able to address the problem of diversity in ranking very efficiently.

Experiments:

In this section, we apply our MRSP algorithm to a couple of real applications: update summarization and query recommendation. As described in Section 2.3, update summarization aims to select sentences conveying the most *relevant, important, diverse*, and *novel* information from the later document set to compose a short summary, given a specific topic and two chronologically ordered document sets. Note that novelty in summarization can be treated as a special kind of diversity, which emphasize the difference between current documents and historical documents. Query recommendation aims to provide diverse and highly related query candidates to cover multiple potential search intents of users and attract more clicks over recommendation. Both of the applications need a ranking method to address diversity, relevance and importance simultaneously. Experiments conducted on these real applications can help demonstrate the effectiveness of our approach on balancing the three goals in ranking.

4.1 Baseline Methods:

For evaluation, we compare our approach with three baseline methods.

• *Baseline-MR (*Zhu, X., *et al*., 2010***):***
Baseline-MR is an extension of the method proposed in (Wen, J.R., 2001). In has two major steps: a) a traditional manifold ranking strategy as described in section 2.1 is applied; b) an additional greedy algorithm is then employed to penalize similar objects.

• *Baseline-MMR (Carbonell, J. and J. Goldstein, 1998)*:
Baseline-MMR is adapted from MMR, (Carbonell, J. and J. Goldstein, 1998) which measures the relevance and diversity independently and provides a linear combination, called “marginal relevance”, as the metric. The ranking score of each object o is computed as follows:

Where Q denotes the query objects, H denotes historical objects, Sim1 and Sim2 are similarity measurements.

• *Baseline-GH:*
Baseline-GH is another baseline method adapted from GRASSHOPPER, which employs an absorbing random walk process to address diversity in ranking. In Baseline-GH, objects are selected iteratively and objects selected so far become absorbing states. The first object is selected according to the personalized Page Rank score. The rest objects are selected according to another metric, i.e. the expected number of visits before absorption. The expected number of visits before absorption can be calculated based on the *fundamental matrix (*Doyle, P.G., 1984*).*
$M = (I - Q)^{-1}$,
Where Q is the sub-matrix of the personalized transition matrix Note $P = \begin{pmatrix} I_G & 0 \\ R & Q \end{pmatrix}$ here G denotes the set of objects selected so far (i.e. absorbed) and I_G denotes the identical matrix with its dimension as the size of G.

4.2 Update Summarization:

4.2.1 Datasets:

Update summarization has been one of the main tasks in TAC2008 and TAC2009 conferences held by

NIST4. They have devoted a lot of manual labor to create the benchmark data for update summarization tasks. TAC2008 has 48 topics and TAC2009 has 44 topics. Each topic is composed of 20 relevant documents from the AQUAINT-2 collection of news articles, and the documents are divided into 2 datasets: Document Set A and Document Set B. Each document set has 10 documents, and all the documents in set A chronologically precede the documents in set B. For update summarization, a 100-word summary is required to be generated for document set B assuming the user has already read the content of set A. We preprocessed the document datasets by removing stop words from each sentence and stemming the remaining words using the Porter's stemmer[5]. For evaluation, four reference summaries generated by human judges for each topic were provided by NIST as ground truth. A brief summary over the two datasets is shown in Table 1.

TABLE 1
Summary of Datasets from TAC2008 and TAC2009

	TAC2008	TAC2009
Track/Task	3/1	3/1
Number of Docs	960	880
Number of Topics	48	44
Ave. Sent. Cnt. per Doc.	21.9	22.8
Ave. Word Cnt. per Sent.	22.6	23.1
Data Sources	AQUAINT-2	AQUAINT-2
Maximum Sum. Length	100 words	100 words

4.2.2 The Benefits of Sink Points:

Our approach can significantly outperform Baseline- MR (p-value<0.05), which also utilizes a manifold ranking approach based on sentence manifold in essentials. As aforementioned, the major difference between the two approaches is that the Baseline-MR method employs an additional greedy algorithm to address novelty and diversity, while our approach introduces sink points into manifold to optimize relevance, importance, diversity, and novelty in one unified process. Here we made some further analysis on these two approaches to show the benefits of sink points based approach.

Here we first compare the novelty and diversity in the summary generated by the two approaches. We use *Obsolete Similarity* (i.e., average similarity between the summary sentences and set A) to measure novelty and *Inter-Sentence Similarity* (i.e., average similarity among the summary sentences) to measure diversity. A lower Obsolete Similarity indicates better novelty and a lower Inter-Sentence Similarity indicates better diversity.

Figures 3(a)~(d) show the average accumulated results of the two measures as the sentences are selected one by one into a summary under the two methods on TAC2008 and TAC2009. Note here we show the accumulated results up to 5 sentences since most summaries generated by the two approaches are within this length. We can see that our approach (using sink points) can consistently obtain lower Obsolete Similarity and Inter-Sentence Similarity during the summarization generation process than Baseline-MR. It demonstrates that by introducing sink points into sentence manifold which can utilize the intrinsic manifold structure, we can better capture both novelty and diversity for update summarization.

Conclusion:

In this paper, we propose a novel MRSP approach to address assortment as well as relevance and significance in ranking. MRSP uses a diverse ranking process over the data manifold, which can naturally find the most relevant and significant objects. In the meantime, by spinning ranked objects into sink points on data assorted, MRSP can effectively prevent redundant objects from receiving a high rank. The integrated MSRP approach can achieve relevance, importance, diversity, and novelty in a unified process. Experiments on tasks of update summarization and query recommendation present strong empirical performance of MRSP.

Experiments for update summarization show that MRSP can achieve comparable performance to the existing best performing systems in TAC antagonisms and outperform other baseline methods. Experiments for query recommendation also determine that our approach can effectively generate diverse and highly related query recommendations.

1) Compute the similarity values $sim(x_i, x_j)$ of each pair of data objects x_i and x_j.
2) Connect any two objects with an edge if their similarity value exceeds 0. We define the affinity matrix W by $W_{ij} = sim(x_i, x_j)$ if there is an edge linking x_i and x_j. Let $W_{ii} = 0$ to avoid self-loops in the graph.
3) Symmetrically normalize W by $S = D^{-1/2}WD^{-1/2}$ in which D is the diagonal matrix with (i,i)-element equal to the sum of the i_{th} row of W.
4) Compute $\Omega = (I - \alpha S)^{-1}$, where $0 \leq \alpha < 1$.
5) Obtain the sub-matrices $\Omega_{11}, \Omega_{12}, \Omega_{21}, \Omega_{22}$ from Ω based on the free points and query points, and the corresponding trimmed vectors y_2.
6) Compute $f^* = \Omega_{22}y_2 - \Omega_{21}\Omega_{11}^{-1}(\Omega_{12}y_2)$.
7) Mark the object x_m with maximum score f_m^* as a new sink point.
8) If the pre-defined number of sink points K is not reached, go to step 5.
9) Return the sink points in the order that they get marked as sink points.

Fig. 2. The Refined MRSP Algorithm.

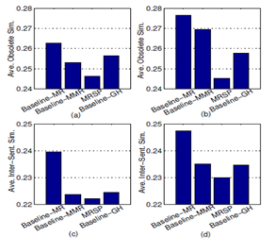

Fig. 3: (a) Average Obsolete Similarity on TAC2008. (b)Average Obsolete Similarity on TAC2009. (c) Average Inter-Sentence Similarity on TAC2008. (d) Average Inter-Sentence Similarity on TAC2009.

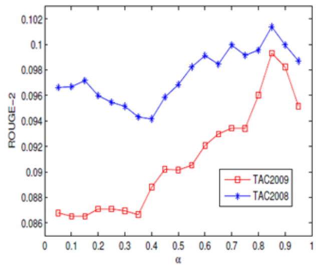

Fig. 4: ROUGE-2 Score vs. Parameter α on MRSP.

REFERENCES

Agrawal, R., S. Gollapudi, A. Halverson and S. Ieong, 2009. Di-versifying search results. In *roceedings of the Second ACM International Conference on Web Search and Data Mining*, WSDM '09, pages 5–14, New York, NY, USA. ACM.

Allan, J., R. Gupta and V. Khandelwal, 2001. Temporal summaries of news topics. In *SIGIR '01: Proceedings of the 24th annual international ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 10–18, New York, NY, USA, ACM.

Beeferman, D. and A. Berger, 2000. Agglomerative clustering of a search engine query log. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, 407–416.

Boldi, P., F. Bonchi, C. Castillo, D. Donato and S. Vigna, 2009. Query suggestions using query-flow graphs. In *Proceedings of the 2009 workshop on Web Search Click Data*, WSCD '09, pages 56–63. New York, NY, USA. ACM.

Boudin, F., M. El-B`eze and J.M. Torres-Moreno, 2008. A scalable MMR approach to sentence scoring for multi-document up-date summarization. In *Coling: Companion volume: Posters*, 23-26. Manchester, UK.

Carbonell, J. and J. Goldstein, 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *SIGIR '98: Proceedings of the 21st annual inter-national ACM SIGIR conference on Research and development in information retrieval*, 335-336. New York, NY, USA.

Clarke, C.L., M. Kolla, G.V. Cormack, O. Vechtomova, A. Ashkan, S.B uttcher and I. MacKinnon, 2008. Novelty and diversity in information retrieval evaluation. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, 659-666.

Dang, H.T. and K. Owczarzak, 2009. Overview of the tac summarization track (draft). In *In Proceedings of the Second Text Analysis Conference (TAC2009)*.

Doyle, P.G. and Snell, J. Laurie (James Laurie), 1984. *Random walks and electric networks / by Peter G. Doyle, J. Laurie Snell*. [Washington, D.C.] : Mathematical Association of America, Includes index.

Du, P., J. Guo, J. Zhang and X. Cheng, 2010. Manifold ranking with sink points for update summarization. In *CIKM '10: Proceeding of the 19th ACM conference on Information and knowledge man-agement*, Toronto, Canada. ACM.

Erkan, G. and D.R. Radev, 2004. Lexrank: graph-based lexical centrality as salience in text summarization. *J. Artif. Int. Res.*, 22(1): 457-479.

Haveliwala, T.H., 2002. Topic-sensitive pagerank. In *Proceedings of the 11th international conference on World Wide Web*, WWW '02, 517-526. New York, NY, USA, ACM.

J¨arvelin, K. and J. Kek¨al¨ainen, 2002. Cumulated gain-based evalu-ation of ir techniques. *ACM Trans. Inf. Syst.*, 20(4): 422-446.

Knight, K. and D. Marcu, 2000. Statistics-based summarization - step one: Sentence compression. In *Proceedings of the Seventeenth Na-tional Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, 703-710. AAAI Press / The MIT Press.

Lan, Y., T.Y. Liu, Z. Ma and H. Li, 2009. Generalization analysis of listwise learning-to-rank algorithms. In *ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning*, pages 577–584, New York, NY, USA, ACM.

Li, L., Z. Yang, L. Liu and M. Kitsuregawa, 2008. Query-url bipartite based approach to personalized query recommendation. In *Proceedings of the 23rd national conference on Artificial intelligence*, 1189–1194.

Li, W., F. Wei, Q. Lu and Y. He, 2008. PNR2: Ranking sentences with positive and negative reinforcement for query-oriented update summarization. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 489–496, Manchester, UK.

Lin, C.Y., 2004. ROUGE: A package for automatic evaluation of summaries. In S. S. Marie-Francine Moens, editor, *Text Sum-marization Branches Out: Proceedings of the ACL-04 Workshop*, 74-81, Barcelona, Spain.

Lin, C.Y. and E. Hovy, 2002. Manual and automatic evaluation of summaries. In *Proceedings of the ACL-02 Workshop on Automatic Summarization*, 45-51. Morristown, NJ, USA.

Mei, Q., J. Guo and D. Radev, 2010.DivRank: the interplay of prestige and diversity in information networks. In *KDD '10: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1009-1018, New York, NY, USA, ACM.

Mei, Q., D. Zhou and K. Church, 2008. Query suggestion using hitting time. In *Proceeding of the 17th ACM conference on Information and knowledge management*, 469-477.

Mihalcea, R. and P. Tarau, 2004. TextRank: Bringing order into texts. In D. Lin and D. Wu, editors, *Proceedings of EMNLP*, pages 404–411, Barcelona, Spain.

Otterbacher, J., G. Erkan and D. Radev, 2005. Using random walks for question-focused sentence retrieval. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 915– 922, Vancouver, British Columbia, Canada.

Radev, D.R., H. Jing and M. Budzikowska, 2000. Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. In *NAACL-ANLP Workshop on Automatic summarization*, pages 21–30, Mor-ristown, NJ, USA.

Radev, D.R. and K.R. McKeown, 1998. Generating natural language summaries from multiple on-line sources. *Comput. Linguist.*, 24: 470-500.

Roweis, S. and L. Saul, 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290: 2323–2326.

Santos, R.L., C. Macdonald and I. Ounis, 2010. Exploiting query reformulations for web search result diversification. In *Pro-ceedings of the 19th international conference on World wide web*, WWW '10, pages 881–890, New York, NY, USA. ACM.

Wan, X., J. Yang and J. Xiao, 2007. Manifold-ranking based topic focused multi-document summarization. In *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 2903–2908, Hyderabad, India, 6-12.

Wei, F., W. Li, Q. Lu and Y. He, 2008. Query-sensitive mutual reinforcement chain and its application in query-oriented multi document summarization. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research anddevelopment in information retrieval*, pages 283–290, New York, NY, USA. ACM.

Wen, J.R., J.Y. Nie and H.J. Zhang, 2001. Clustering user queries of a search engine. In *Proceedings of the 10th international conference on World Wide Web*, pages 162–168.

Zhai, C.X., W.W. Cohen and J. Lafferty, 2003. Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 10–17, New York, NY, USA. ACM.

Zhang, B., H. Li, Y. Liu, L. Ji, W. Xi, W. Fan, Z. Chen and W.Y. Ma, 2005. Improving web search results using affinity graph. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 504–511, New York, NY, USA, ACM.

Zhang, J., X. Cheng, G. Wu and H. Xu, 2008. AdaSum: an adaptive model for summarization. In *CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management*, pages 901–910, New York, NY, USA, ACM.

Zhou, D., O. Bousquet, T.N. Lal, J. Weston and B. Sch¨olkopf, 2004. Learning with local and global consistency. In S. Thrun, L. Saul, and B. Sch¨olkopf, editors, *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA.

Zhou, D., J.Weston, A. Gretton, O. Bousquet and B. Sch¨olkopf, 2004. Ranking on data manifolds. In S. Thrun, L. Saul, and B. Sch¨olkopf, editors, *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA.

Zhu, X., A. Goldberg, J. Van Gael and D. Andrzejewski, 2007. Improving diversity in ranking using absorbing random walks. In *Human Language Technologies: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pages 97–104, Rochester, New York.

Zhu, X., J. Guo and X. Cheng, 2010. Recommending diverse and relevant queries with a manifold ranking based approach. In *SIGIR '10: Workshop on Query Representation and Understanding*, Geneva, Switzerland. ACM.