Ranking on Data Manifold with Sink Points

Xue-Qi Cheng, *Member*, *IEEE*, Pan Du, *Student Member*, *IEEE*,
Jiafeng Guo, Xiaofei Zhu, and Yixin Chen, *Senior Member*, *IEEE*

Abstract—Ranking is an important problem in various applications, such as Information Retrieval (IR), natural language processing, computational biology, and social sciences. Many ranking approaches have been proposed to rank objects according to their degrees of relevance or importance. Beyond these two goals, diversity has also been recognized as a crucial criterion in ranking. Top ranked results are expected to convey as little redundant information as possible, and cover as many aspects as possible. However, existing ranking approaches either take no account of diversity, or handle it separately with some heuristics. In this paper, we introduce a novel approach, Manifold Ranking with Sink Points (MRSPs), to address diversity as well as relevance and importance in ranking. Specifically, our approach uses a manifold ranking process over the data manifold, which can naturally find the most relevant and important data objects. Meanwhile, by turning ranked objects into sink points on data manifold, we can effectively prevent redundant objects from receiving a high rank. MRSP not only shows a nice convergence property, but also has an interesting and satisfying optimization explanation. We applied MRSP on two application tasks, update summarization and query recommendation, 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.

Index Terms—Diversity in ranking, manifold ranking with sink points, update summarization, query recommendation

1 Introduction

Ranking has abundant applications in information retrieval (IR), data mining, and natural language processing. In many real scenarios, the ranking problem is defined as follows: given a group of data objects, a ranking model (function) sorts the objects in the group according to their degrees of relevance, importance, or preferences [15]. For example, in IR, the "group" corresponds to a query, and "objects" correspond to documents associated with the query. However, a mass of relevant objects may contain highly redundant, even duplicated information, which is undesirable for users. Furthermore, the user's needs might be multifaceted or ambiguous. The redundance in top ranked results will reduce the chance to satisfy different users. For example, given a query "zeppelin," if the top ranked search results were all similar articles about the "Zeppelin iPod speaker," it would be a waste of the output space and largely degrade users' search experience even though the results are all highly relevant 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 convey as little redundant information as possible, and cover as many aspects as possible. In this way, we are able to minimize the risk that the information need of the user will not be satisfied. Many real application tasks demand diversity in ranking. For example, in query recommendation, the recommended queries should capture different query intents of different 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.

The issue of diversity in ranking has been widely studied recently. Researchers from various domains have proposed many approaches to address this problem, such as Maximum Marginal Relevance (MMR) [6], subtopic diversity [33], [27], cluster-based centroids selecting [24], categorization-based approach [1], and many other redundancy penalty approaches [34], [17], [30]. However, these methods often treat relevance and diversity separately in the ranking algorithm, sometimes with additional heuristic procedures.

In this paper, we propose a novel approach, named *Manifold Ranking with Sink Points* (**MRSP**), to address diversity as well as relevance and importance in a unified way. Specifically, our approach uses a manifold ranking process [36], [37] over data manifold, which can help find the most relevant and important data objects. Meanwhile, we introduce into the manifold sink points, which are objects whose ranking scores are fixed at the minimum score (zero in our case) during the manifold ranking process. This way, the ranking scores of other objects close to the sink points (i.e., objects sharing similar information with the sink points) will be naturally penalized during the ranking process based on the intrinsic manifold. By turning ranked objects into sink points in the data manifold, we can effectively prevent redundant objects from receiving a high rank. As a result, we

- X.-Q. Cheng, P. Du, J. Guo, and X. Zhu are with the Institute of Computing Technology, Chinese Academy of Sciences, No. 6 Kexueyuan South Road Zhongguancun, Haidian District, Beijing 100190, China. E-mail: cxq@ict.ac.cn, {dupan, guojiafeng, zhuxiaofei}@software.ict.ac.cn.
- Y. Chen is with the Washington University, 1045, One Brookings Drive, St. Louis, MO 63130. E-mail: chen@cse.wustl.edu.

Manuscript received 11 Nov. 2010; revised 9 Aug. 2011; accepted 14 Aug. 2011; published online 21 Sept. 2011.
Recommended for acceptance by G. Cormode.
For information on obtaining reprints of this article, please send e-mail to: tkde@computer.org, and reference IEEECS Log Number TKDE-2010-11-0595.
Digital Object Identifier no. 10.1109/TKDE.2011.190.

1041-4347/13/$31.00 © 2013 IEEE Published by the IEEE Computer Society

can capture diversity as well as relevance and importance during the ranking process. Our proposed approach MRSP has not only a nice convergence property, but also a satisfying optimization explanation.

We applied MRSP to two application tasks, update summarization [10] and query recommendation [39]. Update summarization aims to summarize up-to-date information contained in the new document set given a past document set. The task of query recommendation is to provide alternative queries to help users search and also improve the usability of search engines. In both tasks, diversity is of great concern.

We conducted extensive experiments on the above two tasks. Experiments on update summarization were conducted based on the benchmark data sets of TAC[1] 2008 and TAC 2009. The ROUGE[2] evaluation results show that our approach can achieve comparable performance to the best performing systems in TAC and outperform other baseline methods. Experiments on query recommendation were conducted based on the data set of Microsoft 2006 RFP.[3] Empirical results also demonstrate that our approach is effective in generating highly diverse and highly relevant query recommendations.

The rest of the paper is organized as follows. Section 2 discusses some background and related work. Section 3 describes our approach in detail. Sections 4 and 5 present algorithms and results on update summarization and query recommendation, respectively. The conclusion is made in Section 6.

2 Related Work

2.1 Ranking on Data Manifolds

Ranking on data manifolds is proposed by Zhou et al. [37]. In their approach, data objects are assumed to be points sampled from a low-dimensional manifold embedded in a high-dimensional euclidean space (ambient space). Hereafter, object and point will not be discriminated unless otherwise specified. Manifold ranking is then to rank the data points with respect to the intrinsic global manifold structure [28], [26] given a set of query points.

The manifold ranking algorithm is proposed based on the following two key assumptions: 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 description of the ranking algorithm is described as follows: a weighted network is constructed first, where nodes represent all the data and query points, and an edge is put between two nodes if they are "close." Query nodes are then initiated with a positive ranking score, while the nodes to be ranked are assigned with a zero initial score. All the nodes then propagate their ranking scores to their neighbors via the weighted network. The propagation process is repeated until a global stable state is achieved, and all the nodes except the queries are ranked according to their final scores. The detailed ranking algorithm can be found in [37].

Manifold ranking gives high ranks to nodes that are close to the queries on the manifold (which reflects high relevance) and that have strong centrality (which reflects high importance). Therefore, relevance and importance are well balanced in manifold ranking, similar to Personalized PageRank [12]. However, diversity is not considered in manifold ranking.

2.2 Diversity in Ranking

Beyond relevance and importance, diversity has also been recognized as a crucial criterion in ranking recently [24], [33], [34], [17], [30]. Among the existing work, a well-known approach on introducing diversity in ranking is MMR [6], which constructs a ranking metric combining the criteria of relevance and diversity, but leaving importance unconsidered. Grasshopper [38] 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 DivRank [20], which uses a vertex-reinforced random 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, importance and diversity simultaneously in a unified way is still far from being well resolved.

In [10] and [39], we introduced a novel MRSP algorithm to achieve diversity in ranking in a couple of applications. In this paper, we further extend our research work on MRSP in the following ways. First, we verify that our ranking approach is optimal under the constraints of sink points, local and global consistency. Second, we describe how to reduce the computational cost of the MRSP algorithm. Finally, we conduct extensive experimental analysis to justify the effectiveness and efficiency of our approach.

2.3 Update Summarization

Update summarization is a temporal extension of topic-focused multidocument summarization [23], [31], [35], by focusing on summarizing up-to-date information contained in the new document set given a past document set. There are mainly two kinds of approaches for update summarization, one is abstractive summarization [25], [14], in which some deep natural language processing techniques are leveraged to compress sentences or to reorganize phrases to produce a summary of the text. Another one is extractive summarization [19], [11], [22]. In the extractive approach, update summarization is reduced to a sentence ranking problem, which composes a summary by extracting the most representative sentences from target document set. There are four goals that update summarization aims to achieve:

- *Relevance.* The summary must stick to the topic users are interested in.
- *Importance.* The summary has to neglect trivial content and keep as much important information as possible.
- *Diversity.* The summary should contain less redundant information and cover as many aspects as possible about the topic.
- *Novelty.* The summary needs to focus on the new information conveyed by the later data set as compared with the earlier one under that topic.

[1] Text Analysis Conference: http://www.nist.gov/tac.

[2] http://haydn.isi.edu/ROUGE/.

[3] http://research.microsoft.com/users/nickcr/wscd09.

Technically, *novelty* can be considered as a special kind of diversity since it focuses on the difference between sentences of new documents and those of earlier documents, while *diversity* focuses on the difference between sentences selected already and those to be selected next.

Many approaches have been proposed for update summarization [2], [5], [17], [29]. Boudin et al. [5] described a scalable sentence scoring method, SMMR derived from MMR, where candidate sentences were selected according to a combined criterion of query relevance and sentence dissimilarity. However, neither MMR nor SMMR took the criterion of importance into consideration. Wan [29] presented the TimedTextRank algorithm, a PageRank variation with a time factor, to select new and important sentences for update summarization. They achieved diversity through an additional penalty step based on cosine similarity measurement. Li et al. [17] presented a reinforcement ranking strategy PNR^2 to capture novelty for update summarization. They also penalized redundancy in a similar way as [29] to encourage diversity. It is hard to address the four goals of update summarization in a unified way.

2.4 Query Recommendation

Query recommendation aims to provide alternative queries to help users search and also improve the usability of search engines. It has been employed as a core utility by many industrial search engines. Most of the work on query recommendation focuses on measures of query similarity, where query log data has been widely used in these approaches. For example, Beeferman and Berger [3] applied agglomerative clustering to the click-through bipartite graph to identify related queries for recommendation. Wen et al. [32] proposed to combine both user click-through data and query content information to determine query similarity.

As we can see, most previous work only focuses on the relevance of recommendations, but does not explicitly address the problem of diversity. Mei et al. [21] tackled this problem using a hitting time approach based on the Query-URL bipartite graph. Their approach can recommend more diverse queries by boosting long tail queries. However, long tail queries recommended to users may not be familiar to them, and experimental results [39] show that their approach can sacrifice relevance considerably when improving the diversity.

3 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 [36], [37] to address relevance and importance.

Meanwhile, to address the diversity in ranking, we first introduce the concept of sink points into the data manifold. The sink points are data objects whose ranking scores are fixed at the minimum score (zero in our case) during the ranking process. Hence, the sink points will never spread any ranking score to their neighbors. Intuitively, we can imagine the sink points as the "black holes" on the manifold, where ranking scores spreading to them will be *absorbed* and no ranking scores would *escape* from them.

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 [38]. 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 data set with 100 points as shown in Fig. 1a. There are roughly three groups with different densities. We then connected any two points with the euclidean Distance less than a threshold to form a manifold structure as shown in Fig. 1b. Besides, we randomly selected one data point as a given query, denoted by x_0. We ran the MRSP algorithm over the toy data under different conditions to obtain a series of results as shown in Fig. 1c, 1d, 1e, and 1f, where the vertical axis denotes the ranking score in log scale.

Fig. 1c shows the results when there is no prior preference on any data point (i.e., set $y_i = 1/100$ for each point x_i). This way, the stationary distribution of ranking scores reflects the importance of each point (similar to PageRank). By taking x_0 into account (i.e., set $y_0 = 1$ for point x_0 and $y_i = 0$ otherwise), we obtain a distribution of ranking scores in Fig. 1d. We can then rank the points by their ranking scores, and the top list would be dominated by points from the right group in Fig. 1d. This helps us select relevant and important points given query x_0. Note that in both of the previous two cases, MRSP degrades to the traditional manifold ranking since there are no sink points.

We then selected the top ranked point x_1 (i.e., the most relevant and important point in the right group given x_0), turned it into a sink point, and ran our MRSP algorithm again. We obtained the results in Fig. 1e. As we can see, by turning x_1 into a sink point, our algorithm can well penalize the points close to it in the right group. Thus, points in the middle group were boosted up. Similarly, if we turn the new top ranked point x_2 into a sink point, we will penalize its nearby points in the middle group and make the points in the left group surface, as shown in Fig. 1f. These results show that sink points can work well in the ranking process to penalize nearby points based on the intrinsic manifold structure, and MRSP can address diversity as well as relevance and importance in a unified fashion.

3.3 The Algorithm and Its Convergence

We now describe our MRSP algorithm in detail. Let $\chi = \chi_q \cup \chi_s \cup \chi_r \subset \mathbb{R}^m$ denote a set of data points over the manifold, where $\chi_q = \{x_1, \ldots, x_q\}$ denotes a set of query

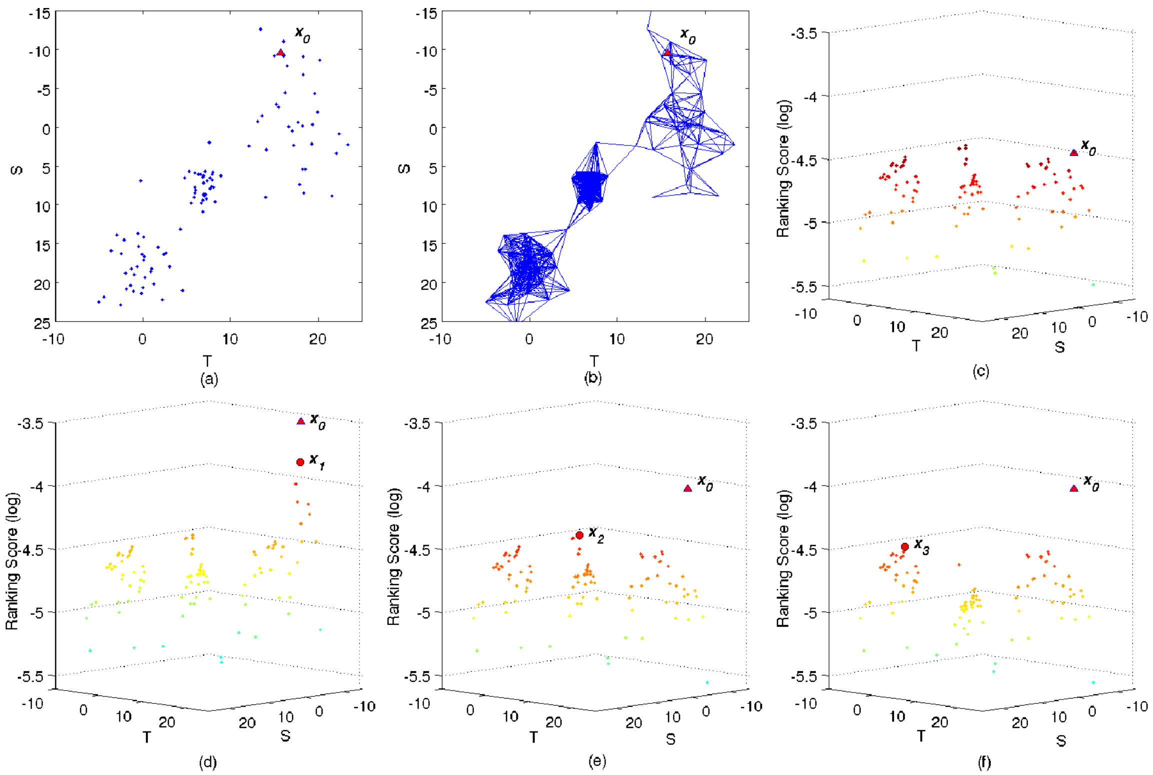


Fig. 1. (a) A data set. (b) The connected weighted network. (c) Ranking score distribution when no prior knowledge on any point. (d) Ranking score distribution given Topic x_0. (e) Ranking score distribution given Topic x_0 and sink point x_1. (f) Ranking score distribution given Topic x_0 and sink points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

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 χ_s as empty.
2. Form the affinity matrix W for the data manifold, where $W_{ij} = sim(x_i, x_j)$ if there is an edge linking x_i and x_j. Note that $sim(x_i, x_j)$ is the similarity between objects x_i and x_j.
3. Symmetrically normalize W as $S = D^{-1/2}WD^{-1/2}$ in which D is a diagonal matrix with its (i, i)-element equal to the sum of the ith row of W.
4. Repeat the following steps if $|\chi_s| < K$:
   a. Iterate $f(t+1) = \alpha S I_f f(t) + (1-\alpha)y$ until convergence, where $0 \leq \alpha < 1$, and I_f is an indicator matrix which is a diagonal matrix with its (i, i)-element equal to 0 if $x_i \in \chi_s$ and 1 otherwise.
   b. Let f_i^* denote the limit of the sequence $\{f_i(t)\}$. Rank points $x_i \in \chi_r$ 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 χ_r to χ_s.
5. Return the sink points in the order that they were selected into χ_s from χ_r.

As we can see, the major difference between MRSP and the traditional manifold ranking algorithm is the introduction of sink points, which in turn affect the ranking process as shown in step $4(a)$-(c). In step $4(a)$, an indicator matrix I_f is used to fix the ranking scores of sink points at zero. As a result, the sink points will not spread any ranking score to their neighbors. We show that the new algorithm with the indicator matrix still achieves convergence.

Theorem 1. *For any fixed sink point set χ_s and hence fixed I_f, the sequence $\{f(t)\}$ in step 3(a) converges to*

$$f^* = (1-\alpha)(I - \alpha S I_f)^{-1} y. \tag{1}$$

Proof. According to the iteration equation, for $t \in \mathbb{N}$,

$$f(t+1) = \alpha S I_f f(t) + (1-\alpha)y.$$

We have

$$f(t) = (\alpha S I_f)^{t-1} f(0) + (1-\alpha) \sum_{i=0}^{t-1} (\alpha S I_f)^i y. \tag{2}$$

Let $\tilde{P} = D^{-1} W I_f$, $\tilde{P}$ is the similarity transformation of $S I_f$ as follows:

$$\begin{aligned} S I_f &= D^{1/2} D^{-1} W D^{-1/2} D^{1/2} I_f D^{-1/2} \\ &= D^{1/2} D^{-1} W I_f D^{-1/2} \\ &= D^{1/2} \tilde{P} D^{-1/2}. \end{aligned}$$

Hence, $\tilde{P}$ and $S I_f$ have the same eigenvalues.

Note that $|\tilde{P}_{ii}| = 0$, according to the Gershgorin circle theorem, we have

$$|\lambda| \leq \sum_{j\neq i} |\tilde{P}_{ij}| \leq 1,$$

where λ is the largest eigenvalue of $\tilde{P}$. Therefore, all the eigenvalues of SI_f are no more than 1.

Since $0 \leq \alpha < 1$, and any eigenvalue of SI_f is no more than 1, we have

$$\lim_{t\to\infty} \left(\alpha SI_f\right)^t = 0,$$

and

$$\lim_{t\to\infty} \sum_{i=0}^{t-1} \left(\alpha SI_f\right)^i = \left(I - \alpha SI_f\right)^{-1}.$$

Hence, from (2), we have

$$f^* = (1-\alpha)(I - \alpha SI_f)^{-1} y.$$

□

We can use this closed form to compute the ranking scores of sentences directly. In large-scale real-world problems, however, an iterative algorithm is preferred for computational reasons.

Since we mainly concern about the ranking scores of the free points, we can further simplify the calculation in MRSP algorithm. The original normalized matrix S in step 2 can be reorganized as a block matrix

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \tag{3}$$

where S_{11} records the relationships between sink points and S_{22} records the relationships between all the query points and free points. The iteration equation in algorithm step 4(a) can then be written as

$$\begin{aligned} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}_{t+1} &= \alpha \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & I_2 \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}_t + (1-\alpha) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= \alpha \begin{pmatrix} 0 & S_{12} \\ 0 & S_{22} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}_t + (1-\alpha) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \end{aligned}$$

where I_2 is an identity matrix, f_1 and f_2 denote the ranking scores of the sink points and others, respectively, and y_1 and y_2 denote the prior on the sink points and others, respectively. Since we only care about the ranking scores of the free points, we only need to compute f_2 with the iteration equation

$$f_2(t+1) = \alpha S_{22} f_2(t) + (1-\alpha) y_2.$$

We then have

Theorem 2. *The sequence* $\{f_2(t)\}$ *converges to*

$$f_2^* = (1-\alpha)(I - \alpha S_{22})^{-1} y_2. \tag{4}$$

3.4 An Optimization Explanation

Our MRSP algorithm also has an interesting optimization explanation, based on a regularization framework for the above algorithm. Let μ be the regularization parameter, and δ be a vector of indicators ($\delta_i = 0$ if x_i is a sink point and 1 otherwise), the cost function associated with the ranking score vector f is

$$\begin{aligned} \mathcal{Q}(f) =& \frac{1}{2} \left(\sum_{i,j=1}^{n} w_{ij} \left\| \frac{\delta_i}{\sqrt{D_{ii}}} f_i - \frac{\delta_j}{\sqrt{D_{jj}}} f_j \right\|^2 + \mu \sum_{i=1}^{n} \|\delta_i f_i - y_i\|^2 \right) \\ =& \frac{1}{2} \left(\sum_{i,j=1}^{n} w_{ij} \left(\frac{\delta_i^2}{D_{ii}} f_i^2 + \frac{\delta_j^2}{D_{jj}} f_j^2 - 2 \frac{\delta_i \delta_j}{\sqrt{D_{ii} D_{jj}}} f_i f_j \right) \right. \\ & \left. + \mu \sum_{i=1}^{n} \left(\delta_i^2 f_i^2 + \delta_i^2 y_i^2 - 2\delta_i f_i y_i \right) \right). \end{aligned}$$

The first term of the right-hand side in the cost function is the *smoothness constraint*, which means that a good ranking function should not change too much between nearby points. The second term is the *fitting constraint*, which means a good ranking function should not change too much from the initial prior assignment. The tradeoff between these two competing constraints is captured by a positive regularization parameter μ. Note that the fitting constraint contains sink points as well as free points.

Since $D_{ii} = \sum_j w_{ij}$, we have

$$\begin{aligned} \mathcal{Q}(f) =& \frac{1}{2} \left(\sum_i \delta_i^2 f_i^2 + \sum_j \delta_j^2 f_j^2 - 2 \sum_{i,j} w_{ij} \frac{\delta_i \delta_j}{\sqrt{D_{ii} D_{jj}}} f_i f_j \right. \\ & \left. + \mu \sum_{i=1}^{n} \left(\delta_i^2 f_i^2 + \delta_i^2 y_i^2 - 2\delta_i f_i y_i \right) \right) \\ =& (I_f f)^T (I - S) I_f f + \mu (I_f f - y)^T (I_f f - y). \end{aligned}$$

The optimal solution of f then can be obtained by minimizing the cost function $\mathcal{Q}(f)$

$$f^* = \arg\min_{f\in\mathcal{F}} \mathcal{Q}(f).$$

Differentiating $\mathcal{Q}(f)$ with respect to f, we have

$$\frac{\partial \mathcal{Q}}{\partial f}|_{f=f^*} = 2I_f(I-S)I_f f^* + 2\mu I_f (I_f f^* - y) = 0$$

which leads to

$$\begin{aligned} & I_f f^* - I_f S I_f f^* + \mu I_f f^* - \mu I_f y = 0 \\ & I_f[(1+\mu)I - SI_f] f^* = \mu I_f y. \end{aligned}$$

Let $\alpha = 1/(1+\mu), \beta = \mu/(1+\mu)$, we get

$$\begin{aligned} & I_f(I - \alpha SI_f) f^* = \beta I_f y \\ & \begin{pmatrix} O & O \\ O & I_2 \end{pmatrix} \left[\begin{pmatrix} I_1 & O \\ O & I_2 \end{pmatrix} - \alpha \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} O & O \\ O & I_2 \end{pmatrix} \right] \begin{pmatrix} f_1^* \\ f_2^* \end{pmatrix} \\ & = \beta \begin{pmatrix} O & O \\ O & I_2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \end{aligned}$$

which leads to

$$\begin{pmatrix} O & O \\ O & I_2 - \alpha S_{22} \end{pmatrix} \begin{pmatrix} f_1^* \\ f_2^* \end{pmatrix} = \beta \begin{pmatrix} O \\ y_2 \end{pmatrix},$$

from which we get

$$f_2^* = \beta (I - \alpha S_{22})^{-1} y_2,$$

which is exactly the closed form of ranking function for free points as shown in (4).

3.5 Computational Refinement

Thanks to the sink points we introduced, the ranking problem in MRSP can be recognized as a Dirichlet problem. Therefore, the computational efficiency of MRSP can be further improved.

Given

$$f(t+1) = \alpha S I_f f(t) + (1-\alpha)y,$$

and its convergence property, we have the form below as it converges:

$$f^* - \alpha S I_f f^* = (1-\alpha)y.$$

The equation above can be transformed into

$$\begin{aligned}(I_s + I_f)f^* - \alpha S I_f f^* &= (1-\alpha)y \\ (I - \alpha S)I_f f^* &= (1-\alpha)y - I_s f^*,\end{aligned}$$

where $I_s + I_f = I$. Let $\Omega = (I - \alpha S)^{-1}$. We have

$$I_f f^* = (1-\alpha)\Omega y - \Omega I_s f^*.$$

If we use the block matrix form of S in (3) and define the corresponding block matrix of Ω

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix},$$

we have

$$\begin{pmatrix} 0 \\ f_2^* \end{pmatrix} = (1-\alpha)\begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}\begin{pmatrix} f_1^* \\ 0 \end{pmatrix},$$

$$\begin{pmatrix} 0 \\ f_2^* \end{pmatrix} = \begin{pmatrix} (1-\alpha)(\Omega_{11}y_1 + \Omega_{12}y_2) - \Omega_{11}f_1^* \\ (1-\alpha)(\Omega_{21}y_1 + \Omega_{22}y_2) - \Omega_{21}f_1^* \end{pmatrix}.$$

Hence, we get the following equations:

$$\begin{cases} 0 &= (1-\alpha)\Omega_{11}y_1 + (1-\alpha)\Omega_{12}y_2 - \Omega_{11}f_1^*, \\ f_2^* &= (1-\alpha)\Omega_{21}y_1 + (1-\alpha)\Omega_{22}y_2 - \Omega_{21}f_1^*, \end{cases}$$

which implies

$$\begin{cases} \Omega_{11}[(1-\alpha)y_1 - f_1^*] = -(1-\alpha)\Omega_{12}y_2, \\ f_2^* = \Omega_{21}[(1-\alpha)y_1 - f_1^*] + (1-\alpha)\Omega_{22}y_2. \end{cases} \tag{5}$$

By combining the two equations in (5), we get

$$f_2^* = (1-\alpha)(\Omega_{22}y_2 - \Omega_{21}\Omega_{11}^{-1}(\Omega_{21}y_2)). \tag{6}$$

Hence, we can use (6) to calculate the ranking scores of free points instead of (4).

However, once a free point is selected and gets sinked, we need to reorganize the matrix S as in (3) (i.e., to preform a row and column switch). This will result in a recalculation of the matrix $\Omega = (I - \alpha S)^{-1}$, which is computationally expensive. Fortunately, we only need to acquire Ω once before the first iteration, because the reorganization of matrix S has the same effect as the reorganization of matrix Ω, which is shown in Theorem 3.

Theorem 3. *For $i = 1, \ldots, n$, let p_i be a row switching matrix adapted from the identity matrix I, let $P = \prod_{i=1}^{n} p_i$ and $\Omega = (I - \alpha S)^{-1}$, then it holds that*

$$(I - \alpha P S P^T)^{-1} = P \Omega P^T.$$

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.

Proof. Since we know

$$I - \alpha P S P^T = P P^T - P(\alpha S)P^T = P(I - \alpha S)P^T$$

and, $P^T = \prod_{i=n}^{1} p_i = P^{-1}$ and $(P^T)^{-1} = P$, we have

$$\begin{aligned}(I - \alpha P S P^T)^{-1} &= (P(I - \alpha S)P^T)^{-1} \\ &= (P^T)^{-1}(I - \alpha S)^{-1}P^{-1} \\ &= P(I - \alpha S)^{-1}P^T \\ &= P\Omega P^T.\end{aligned} \tag{7}$$

□

Therefore, when a free point needs to be sinked, the overhead of inverse operation contained in the computation of matrix Ω is saved. The new ranking score of free points can be obtained by (6) after the reorganization of matrix Ω.

Let N be the total number of points, K the number of sink points, and Q the number of query points, the cost of computing ranking scores of free points using (6) is $O(NK^2 + NK + NQ + N - K^2 - K)$. Since usually $K \ll N$, $Q \ll N$, the cost of (6) is much lower than that of (4), which is $O(N^3)$. Therefore, ranking free points using (6) in MRSP is much more efficient than using (4). Empirically, MRSP takes about 0.03792 s, on average, to generate a summary, which is significantly more efficient (p-value < 0.01) than Grasshopper approach (~ 0.04463 s), and comparative to MMR approach (~ 0.03570 s) in our experiments.

3.6 The Refined MRSP Algorithm

Based on above analysis, the refined algorithm of MRSP is described in Fig. 2.

Note that in step 5, matrix Ω 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.

4 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* [37]: Baseline-MR is an extension of the method proposed in [30]. It has two major steps: 1) a traditional manifold ranking strategy as described in Section 2.1 is applied; 2) an additional greedy algorithm is then employed to penalize similar objects.
- *Baseline-MMR* [6]: Baseline-MMR is adapted from MMR [6], 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:

$$MMR(o) = \lambda Sim_1(o,Q) - (1-\lambda) \max_{o_h \in H} Sim_2(o, o_h),$$

  where Q denotes the query objects, H denotes historical objects, Sim_1 and Sim_2 are similarity measurements.
- *Baseline-GH*[38]: Baseline-GH is another baseline method adapted from GRASSHOPPER [38], 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 PageRank score. The rest objects are selected according to another metric, i.e., the expected number of visits before absorption [38]. The expected number of visits before absorption can be calculated based on the *fundamental matrix* [9]

$$M = (\boldsymbol{I} - Q)^{-1},$$

  where Q is the submatrix of the personalized transition matrix

$$P = \begin{pmatrix} \boldsymbol{I}_G & \boldsymbol{0} \\ R & Q \end{pmatrix}.$$

  Note here G denotes the set of objects selected so far (i.e., absorbed) and $\boldsymbol{I}_G$ denotes the identical matrix with its dimension as the size of G.

TABLE 1
Summary of Data Sets 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 Update Summarization

4.2.1 Data Sets

Update summarization has been one of the main tasks in TAC2008 and TAC2009 conferences held by NIST.[4] 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 two data sets: 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 data sets 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 data sets is shown in Table 1.

4.2.2 Implementation

In this experiment, we construct the sentence manifold according to the pairwise similarity values $sim(x_i, x_j)$ between sentences x_i and x_j. x_i is a term-vector recording the Term Frequency-Inverse Sentence Frequency *(tf-isf)* values of the sentence. The pairwise similarity is calculated with the standard Cosine measure. Then we connect any two points 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.

[4] http://www.nist.gov.

[5] http://www.tartarus.org/martin/PorterStemmer/.

Then we apply MRSP algorithm on the sentence manifold to generate the update summary. The topic sentence is set as the query point and both a representative sentence from the earlier data set and sentences already selected for summary are turned into sink points during the ranking process. Note that there might be different ways to represent the earlier data set A as an initial sink point, including summary of A, the most representative sentence of A, all the sentences from A, and an aggregated pseudosentence from A. In our experiments, we find that among all these representations, better performance of update summarization can be achieved when adopting the most representative sentence of A or an aggregated pseudosentence from A as the representation of A.

4.2.3 Evaluation Metric

ROUGE [18] has become the most frequently used toolkit for automatic summarization evaluation, as it produces the most reliable scores in correspondence to human evaluations. It measures summary quality by counting the number of overlapping units such as n-gram, word sequences, and word pairs between the computer-generated summary and the ideal summaries created by human. The n-gram recall measure, ROUGE-N, is computed as

$$ROUGE - N = \frac{\sum_{S\in\{Refs\}}\sum_{gram_n\in S} Cnt_{match}(gram_n)}{\sum_{S\in\{Refs\}}\sum_{gram_n\in S} Cnt(gram_n)},$$

where n stands for the length of the n-gram, $Cnt_{match}(gram_n)$ is the maximum number of n-grams co-occurring in a candidate summary and a set of reference summaries $Refs$, and $Cnt(gram_n)$ is the number of n-gram in the reference summaries.

In our evaluation, we use the ROUGE-2 (bigram-based) and ROUGE-SU4 (an extended version of ROUGE-2) automatic metrics, which have been shown to correlate well with human judgments based on comparison with a single model [18]. They were also used as official automatic evaluation metrics for TAC2008 and TAC2009, respectively. The results were obtained using ROUGE version 1.5.5 under the settings used for TAC2008.

4.2.4 Qualitative Comparison

We first show a qualitative comparison to gain some intuition on the differences between the summaries generated by our approach and the baseline methods. We randomly selected one topic from the 48 topics of TAC2008 data set as an example, which is about "*the investigation of Jack Abramoff and others related to lobbying activities.*" We show the four reference summaries, provided by NIST as the ground truth, and the summaries generated by our approach and other three baselines in Table 2.

We manually annotated the important facts covered by the ground truth with circled numbers and identified in total 15 topic-related facts in the references. We then manually annotated these facts in the summaries generated by our approach and other baseline methods. We list the statistics of the fact coverage of MRSP and baseline methods in Table 3. The *Facts* column shows the major related words, the *Weight* column shows the unique times that the corresponding fact has appeared in the four references, and the rest of columns record the fact coverage status of different approaches.

The results in Table 3 show that our MRSP approach covers the most facts as compared to the other baselines. Baseline-MR captures the two most agreed facts (i.e., weight $= 4$) ③ and ④, but no other facts. It indicates that Baseline-MR may not address diversity very well. Baseline-MMR only covers one fact ⑤, which is not among the most important facts. Baseline-GH is a strong baseline, which captures only 1 fact fewer than MRSP. In addition, we observe that every sentence extracted by MRSP covers some topic-related facts. On the contrary, all baseline methods generate sentences covering no facts, which is a waste of the limited summary space.

We further focus on the facts that appeared at least in two references (i.e., weight ≥ 2), which represent some commonly agreed facts in summary given the topic. There are five topics which meet the criterion, named ①, ③, ④, ⑤, and ⑪. We can see that MRSP can cover four out of these five topics. In fact, the reference summary can only cover 3.75 out of these five topics on average. There is only one reference, RefSum IV, that covers more facts than MRSP.

4.2.5 Quantitative Comparison

We also conducted quantitative comparison between MRSP and the baselines. The performance comparison based on update summarization tasks of TAC2008 and TAC2009 is shown in Tables 4 and 5, respectively.

Except for the baselines described in Section 4.1, here we also adopt several other baselines for update summarization, including Baseline-L, Baseline-U, and the best-performing systems in TAC2008 and TAC2009. Baseline-L and Baseline-U are two standard baseline methods provided by NIST on TAC. Baseline-L takes all the leading sentences (up to 100 words) in the most recent document. It provides a lower bound on what can be achieved with those extractive summarizer [8]. Baseline-U generates a summary consisting of sentences that have been manually selected from the data set by a team of five human "sentence-extractors" from the University of Montreal. It provides an approximate upper bound on what can be achieved with a purely extractive summarizer. This baseline method is only available on TAC2009. In Table 4, S14 represents the best performing system in TAC2008. It is also an extractive summarization approach. In Table 5, S34 represents the best performing system in TAC2009. However, since S34 is not a purely extractive summarization approach (with massive abstractive techniques), we also show the best performing extractive summarization approach on TAC2009, denoted as system S24.

From the results on TAC2008 shown in Table 4, we can see that MRSP achieves 70.3 percent improvement on ROUGE-2 and 46.6 percent improvement on ROUGE-SU4, compared with Baseline-L. Our approach also achieves better quality than S14 in terms of both ROUGE-2 and ROUGE-SU4, and significantly outperforms the other baselines (p-value <0.05). Similarly, from the results on TAC2009 shown in Table 5, we can observe that our approach achieves 93.2 percent improvement on ROUGE-2 and 51.5 percent improvement on ROUGE-SU4, compared with Baseline-L. MRSP also obtains comparable performance to S34 and Baseline-U. Besides, our

TABLE 2
Qualitative Comparison on a Topic in TAC2008

Topic	Describe the ***investigation*** of ***Jack Abramoff*** and ***others*** related to ***lobbying activities***.
RefSum I	①*Scanlon pleaded guilty* to conspiring to bribe a congressman and other public officials. His help was expected to be crucial in the Justice Department's wide-ranging Abramoff investigations. On Dec. 28, 2005 the ②*Justice Department was urgently pressing Abramoff* to settle fraud and bribery allegations as the investigation continued. A court date was set for Jan. 9, 2006 as Senate and criminal investigations continued. On Jan. 3 ③*Abramoff pleaded guilty* and ④*agreed to cooperate*. He ④*agreed to provide evidence and testimony* in the Justice Department's probe of corruption in Congress and executive branch agencies.
RefSum II	①*Michael Scanlon pleaded guilty* to conspiring to bribe public officials and ⑤*agreed to pay back* over $19 million to ⑥*Indian tribal clients*. His possible 5-year prison ⑦*sentence could be reduced* should he cooperate with prosecutors. ③*Abramoff pleaded guilty* to federal conspiracy, fraud and tax evasion charges and ④*agreed to cooperate* with investigators. ⑧*He did not identify officials by name*. The next day he ③*pleaded guilty* to fraud and conspiracy in Florida. Panicked ⑨*lawmakers, including President Bush, rushed to distance themselves from Abramoff* by returning campaign money he gave them. ⑩*The money is being returned or redirected to charities*.
RefSum III	③*Abramoff pleaded guilty* to conspiracy, fraud and tax evasion and ④*agreed to cooperate* with federal officials investing political corruption. Abramoff is also facing trial on ⑪*fraud charges* in Florida connected with purchasing *a casino cruise chip company*. In his plea agreement, Abramoff committed to making ⑫*25 million in restitution to his victims* and paying ⑬*1.2 million to the IRS* for tax evasion. ⑭*His prison sentence could be lowered* depending on assistance he provides in the corruption investigation of as many as 20 members of Congress or top congressional aides.
RefSum IV	Abramoff's extensive corruption schemes are the subject of a Senate probe and criminal investigations by the Department of Justice and Florida prosecutors. Investigators are looking at congressmen, aides, government officials, and lobbyists, including Rep. John Doolittle, Stephen Griles, and Tony Rudy, and will follow the case wherever it leads. ①*Scanlon pled guilty* to bribery and ⑤*agreed to make restitution* to tribes. His help would be crucial to the Abramoff investigation. ③*Abramoff pled guilty* to conspiracy, fraud, and tax evasion and ④*agreed to cooperate* with prosecutors. Co-defendant ⑮*Adam Kidan agreed to testify against Abramoff* regarding ⑪*fraud in a ship company purchase.*
Baseline-MMR	Jack Abramoff liked to slip into dialogue from "The Godfather" as he led his lobbying colleagues in planning their next conquest on Capitol Hill. Members of Congress who once counted on super-lobbyist Jack Abramoff to help finance their campaigns have begun ⑤*returning the cash* they got from him and his clients, signalling a growing worry that ethics - and the scandal surrounding Abramoff - will become issues that could affect close House and Senate races in next year's midterm elections. Former lobbyist Jack Abramoff's second guilty plea in two days sealed this role as a star witness in the federal government's largest congressional corruption investigation in decades.
Baseline-GH	A onetime congressional staffer who became a top partner to lobbyist Jack Abramoff ①*pleaded guilty* to conspiring to bribe a congressman and other public officials and ⑤*agreed to pay back* more than 9 million he ⑥*fraudulently charged Indian tribal clients*. Former high-powered lobbyist ③*Jack Abramoff*, a longtime associate of top Republican leaders, *pleaded guilty* Tuesday to conspiracy, fraud and tax evasion and ④*agreed to cooperate* with federal officials investigating political corruption in the nation's capital. Jack Abramoff liked to slip into dialogue from "The Godfather" as he led his lobbying colleagues in planning their next conquest on Capitol Hill.
Baseline-MR	Jack Abramoff liked to slip into dialogue from "The Godfather" as he led his lobbying colleagues in planning their next conquest on Capitol Hill. Former high-powered lobbyist ③*Jack Abramoff*, a longtime associate of top Republican leaders, *pleaded guilty* Tuesday to conspiracy, fraud and tax evasion and ④*agreed to cooperate* with federal officials investigating political corruption in the nation's capital. President George Bush's re-election campaign is returning thousands of dollars in campaign contributions connected to ⑨*Jack Abramoff*, the Republican lobbyist who has admitted to *bribing lawmakers* in a case that is leading to a widespread corruption probe involving top Washington officials.
*MRSP**	Former high-powered lobbyist ③*Jack Abramoff*, a longtime associate of top Republican leaders, *pleaded guilty* Tuesday to conspiracy, fraud and tax evasion and ④*agreed to cooperate* with federal officials investigating political corruption in the nation's capital. President George Bush's re-election campaign is returning thousands of dollars in campaign contributions connected to ⑨*Jack Abramoff*, the Republican lobbyist who has admitted to *bribing lawmakers* in a case that is leading to a widespread corruption probe involving top Washington officials. A onetime ①*congressional staffer* who became a top partner to lobbyist Jack Abramoff *pleaded guilty* to conspiring to bribe a congressman and other public officials and ⑤*agreed to pay back* more than 9 million he ⑥*fraudulently charged Indian tribal clients*.

approach can significantly outperform the other baseline methods (p-value < 0.05). Note that our approach can even outperform the Baseline-U method, which is the approximate upper bound baseline system of TAC2009 provided by NIST.

4.2.6 The Benefits of Sink Points

As we can see from the results in Tables 4 and 5, 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 *Intersentence Similarity* (i.e., average similarity among the summary sentences) to measure diversity. A lower Obsolete Similarity indicates better novelty and a lower Intersentence Similarity indicates better diversity.

TABLE 3
Comparison on Facts Coverage over Different Approaches

	Facts	Weight	Baseline-MMR	Baseline-GH	Baseline-MR	**MRSP***
③	Abramoff pleaded guilty	4	-	√	√	√
④	Abramoff agreed to cooperate	4	-	√	√	√
①	Scanlon pleaded guilty	3	-	√	-	√
⑤	Scanlon agreed to pay back	2	√	√	-	√
⑪	Abramoff fraud in a ship company purchase	2	-	-	-	-
⑨	Lawmakers distance themselves from Abramoff	1	-	-	√	√
⑥	Scanlon fraudulently charged Indian tribal clients	1	-	√	-	√
⑧	Abramoff did not identify officials by name	1	-	-	-	-
②	Justice Department presses Abramoff	1	-	-	-	-
⑦	Scanlon's sentence could be reduced	1	-	-	-	-
⑩	The lawmakers is returning money	1	-	-	-	-
⑫	Abramoff committed to making restitution	1	-	-	-	-
⑬	Abramoff committed to paying tax evasion	1	-	-	-	-
⑭	Abramoff's prison sentence could be lowered	1	-	-	-	-
⑮	Kidan agreed to testify against Abramoff	1	-	-	-	-

TABLE 4
Comparison on TAC2008

	ROUGE-2 (%)	95% C.I.	ROUGE-SU4 (%)	95% C.I.
MRSP	**0.10217 (+70.3%)**	**[0.08909 - 0.11692]**	**0.13778 (+46.6%)**	**[0.12676 - 0.15017]**
S14	0.101 (+68.3%)	[0.08827 - 0.11582]	0.137 (+45.7%)	[0.12546 - 0.14917]
Baseline-GH	0.09603 (+60.1%)	[0.08304 - 0.11082]	0.13301 (+40.5%)	[0.12197 - 0.14602]
Baseline-MMR	0.09287 (+54.8%)	[0.07953 - 0.10610]	0.12996 (+38.3%)	[0.11869 - 0.14141]
Baseline-MR	0.09171 (+52.9%)	[0.07728 - 0.10660]	0.13124 (+39.6%)	[0.11822 - 0.14404]
Baseline-L	0.05859	[0.04952 - 0.06813]	0.09283	[0.08362 - 0.10211]

Numbers in parentheses indicates relative improvement over Baseline-L.

TABLE 5
Comparison on TAC2009

	ROUGE-2 (%)	95% C.I.	ROUGE-SU4 (%)	95% C.I.
S34	0.10386 (+100.4%)	[0.09190 - 0.11580]	0.13851 (+52.4%)	[0.12771 - 0.14920]
MRSP	**0.09932 (+93.2%)**	**[0.08781 - 0.11096]**	**0.13771 (+51.5%)**	**[0.12668 - 0.14850]**
Baseline-U	0.09820 (+91.0%)	[0.08633 - 0.11004]	0.13631 (+50.0%)	[0.12517 - 0.14679]
S24	*0.09615 (+87.0%)*	[0.08417 - 0.10825]	*0.13520 (+48.7%)*	[0.12382 - 0.14681]
Baseline-GH	0.09536 (+85.5%)	[0.08300 - 0.10730]	0.13638 (+50.0%)	[0.12539 - 0.14710]
Baseline-MMR	0.09237 (+80.0%)	[0.08079 - 0.10379]	0.13197 (+45.2%)	[0.12215 - 0.14156]
Baseline-MR	0.09002 (+75.1%)	[0.07860 - 0.10209]	0.12908 (+42.0%)	[0.11831 - 0.14067]
Baseline-L	0.05142	[0.04267 - 0.06000]	0.09091	[0.09289 - 0.09884]

Numbers in parentheses indicates relative improvement over Baseline-L.

Figs. 3a, 3b, 3c, and 3d 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 Intersentence 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.

Meanwhile, we show the average accumulated ROUGE-2 scores in Figs. 3e and 3f. The results show that our MRSP approach can consistently obtain higher ROUGE-2 scores during the summarization generation process than Baseline-MR. In other words, our approach can always select better sentences for summarization as compared with Baseline-MR. The overall results in Fig. 3 demonstrate that by introducing sink points into sentence manifold to simultaneously address the four issues in a unified way, MRSP can achieve better performance on update summarization.

To further demonstrate the effectiveness of the sink points, we also compare MRSP with all the other three baseline methods in terms of the average Obsolete Similarity and average Intersentence Similarity on TAC2008 and TAC2009, as shown in Fig. 4. Clearly, our approach achieves the lowest Obsolete Similarity and Intersentence Similarity on both data sets. The results demonstrate that MRSP performs better than baselines on achieving novelty and diversity.

4.2.7 Parameter Tuning

There is only one parameter α in the MRSP algorithm, which is a balance factor between the influence of the intrinsic manifold structure and the prior knowledge on each sentence. Fig. 5 shows the influence of α on the summarization performance. As we can see, the summarization approach performs not so well when α is small, which may be due to over emphasis on the prior knowledge. However, we can also notice the degradation of performance when α approaches 1, which shows putting too much weight on the

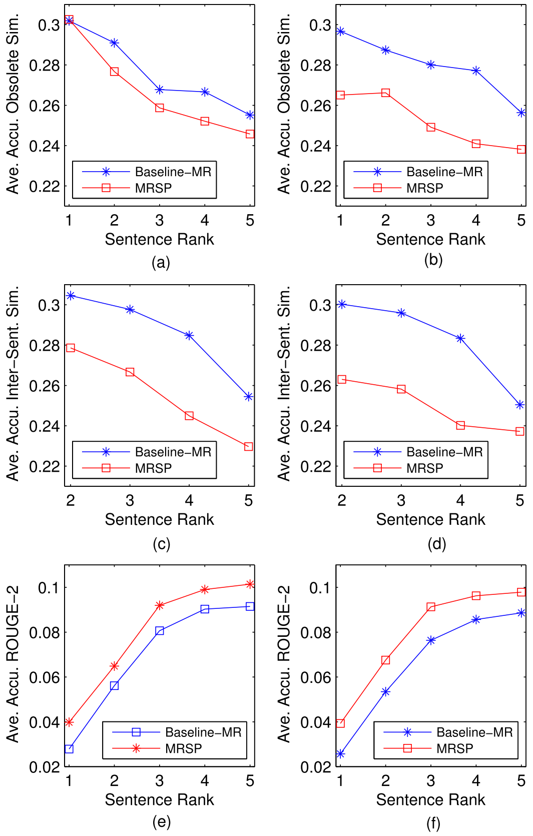


Fig. 3. (a) Average accumulated obsolete similarity on TAC2008. (b) Average accumulated obsolete similarity on TAC2009. (c) Average accumulated intersentence similarity on TAC2008. (d) Average accumulated intersentence similarity on TAC2009. (e) Average accumulated ROUGE-2 on TAC2008. (f) Average accumulated ROUGE-2 on TAC2009.

influence of structure may not work well either. As a result, MRSP achieves the best performance when $\alpha = 0.85$ on both benchmarks of TAC2008 and TAC2009.

4.3 Query Recommendation

4.3.1 Data Set

The experiments are based on the Microsoft 2006 RFP data set[6] which contains about 15 million queries (from US users) that were sampled over one month in May 2006. We clean the raw data by ignoring non-English queries, converting letters into lower case, and trimming each query. To further reduce the noise in clicks, the click-through between a query and a URL with a frequency less than 3 was removed. After cleaning, we get click-through data with 191,585 queries, 251,427 URLs and 318,947 edges. Similar to [4], we randomly sampled 150 queries with frequencies between 700 and 15,000 for evaluation.

To conduct manual evaluation for comparing different recommendation methods, we invite three human judges to label the recommendations in the pool manually. For each query, we create a recommendation pool by merging the topmost (e.g., 10 in our work) recommendations from all the methods. For each test query, the human judges are required to identify relevant recommendations and further group them into clusters according to their search intent. Since the labeling task is costly, we randomly pick 50 queries for manual evaluation.

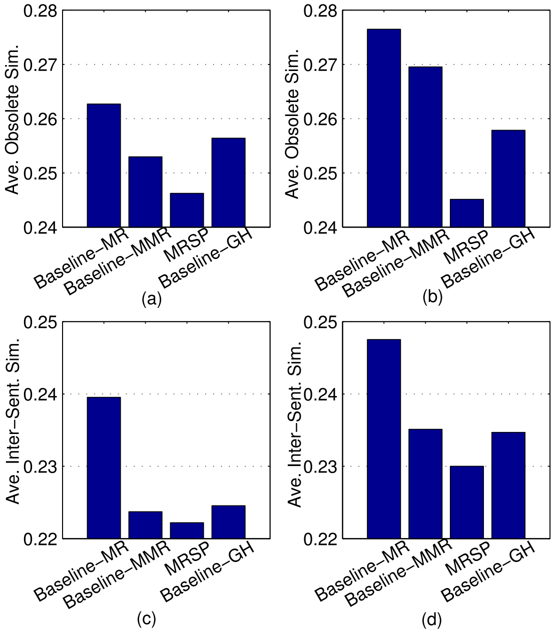


Fig. 4. (a) Average obsolete similarity on TAC2008. (b) Average obsolete similarity on TAC2009. (c) Average intersentence similarity on TAC2008. (d) Average intersentence similarity on TAC2009.

4.3.2 Implementation

In this experiment, we first build the query manifold by identifying and connecting the k-nearest neighbors of each query. Here we leverage the click-through information in query logs to represent the query vector. The basic idea is that if two queries share many clicked URLs, they have similar search intent to each other [16]. Therefore, we model queries in terms of query-URL vectors, instead of query-term vectors. We represent each query q_i as an L_2-normalized vector, where each dimension corresponds to a unique URL in the click-through data. Here k was empirically set to 30 in our experiments. We define an affinity matrix W for the query manifold, where $w_{ij} = exp[-d(q_i, q_j)^2/2\sigma^2]$ if there

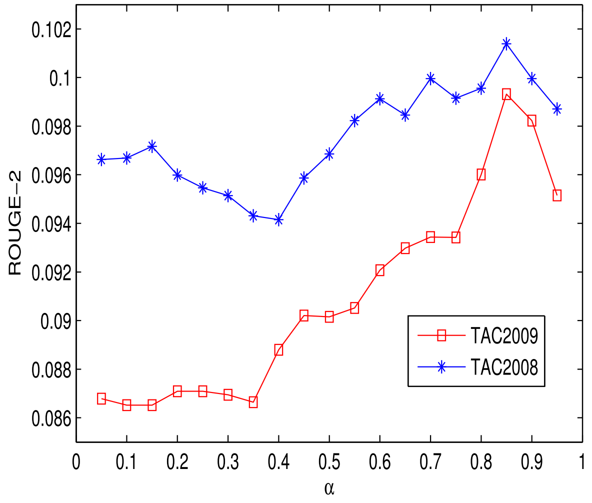


Fig. 5. ROUGE-2 score versus parameter α on MRSP.

[6] http://research.microsoft.com/users/nickcr/wscd09/.

TABLE 6
Examples of Query Recommendations Provided by Different Approaches (Top 10 Results)

query	Baseline-Naive	Baseline-MR	Baseline-MMR	Baseline-GH	MRSP
abc	abc shows	abc tv	abc shows	abc tv	abc tv
	abc television	abc news	abc breaking news	abc news	abc news
	abc tv	abc family	associated builders and contractors	abc family	abc nightline
	abc news	abc shows	abc nightline	abc shows	abc family
	abc breaking news	abc breaking news	abc tv	abc breaking news	associated builders and contractors
	abc family	abc world news	abc television	nightline	abc shows
	abc sports	news stories	abc family	goodmorning america	abc daytime
	abc world news	goodmorning america	abc sports	abc sports	goodmorning america
	world news tonight	abc soaps	abc daytime	abc daytime	abc sports
	abc soap operas	abc sports	goodmorning america	national news	abc soap operas
yamaha	yamaha america	yamaha motors	yamaha america	yamaha motor	yamaha motor
	yamaha motor corp	yamaha marine	yamaha atv parts	yamaha america	yamaha motor corp
	yamaha motor	yamaha atv	yamaha boat motors	yamaha motor corp	yamaha america
	yamaha motor co	yamaha motorcycles	yamaha motor corp	yamaha motorcycles	yamaha marine
	yamaha motorcycle	yamaha motor	yamaha snowmobiles	motorcycles	yamaha atv
	yamaha motors	yamaha moter co	yamaha motor	yamaha marine	yamaha snowmobiles
	yamaha motorcycles	yamaha motor corp	yamaha drums	yamaha atv	yamaha drums
	yamaha quads	yamaha snowmobiles	yamaha guitars	yamaha motorcycle parts	yamaha guitars
	yamaha snowmobiles	yamaha outboard	yamaha motorcycles	yamaha snowmobiles	yamaha quads
	yamaha scooters	yamaha boat motors	yamaha atvs	yamaha quads	yamaha boat motors

is an edge linking q_i and q_j, σ is empirically set to 1.25. We set $w_{ii} = 0$ to avoid self-loops in the graph. Since most queries are irrelevant to the input query, we can use a breadth-first search strategy to construct a submanifold to save the computational cost.

Based on the query manifold, we apply our MRSP algorithm to return the top-K recommended queries. In this task, the input query is set as topic node, and the queries already selected for recommendation are set as sink points during the ranking process. In our experiments, top 10 recommendations are considered for a query and our MRSP method takes about 0.1s, on average, to generate the recommendations.

4.3.3 Evaluation Metrics

With the human labeled data, we evaluate the quality of the recommendations produced by different approaches using the following two measures: 1) *α-normalized Discounted Cumulative Gain (α-nDCG)* [7] which has been widely used in the TREC Web track[7] diversity task; and 2) *Intent-Coverage*. Both of the two metrics range from 0 to 1, with 1 as the best value, and 0 the worst.

α-nDCG. The α-nDCG, which rewards diversity in ranking, is a new version of the nDCG [13], the normalized Discounted Cumulative Gain measure. When $\alpha = 0$, the α-nDCG measure corresponds to the standard nDCG, and when α is closer to 1, the diversity is rewarded more in the metric. The key difference between α-nDCG and nDCG is that they use different gain value. For each recommendation, the gain value $G(k)$ of α-nDCG is defined as

$$G(k) = \sum_{i=1}^{I} J_i(k)(1-\alpha)^{C_i(k-1)}, \tag{8}$$

where $C_i(k-1)$ is the number of relevant recommendations found within the top $k-1$ recommendations for intent i, $J_i(k)$ is a binary variable indicating whether the recommendation at rank k belongs to intent i or not, and I is the total number of unique intents for each test query. The computation of α-nDCG exactly follows the procedure described in [7] with $\alpha = 0.5$.

Intent-coverage. The Intent-Coverage measures the proportion of unique intents covered by the top k recommended queries for each test query. Since each intent represents a specified user information need, higher Intent-Coverage indicates larger probability to satisfy different users. Note that the Intent-Coverage is different from the diversity measure used in automatic evaluation, since only relevant recommendations to the test query will be considered in Intent-Coverage. Therefore, Intent-Coverage can better reflect the diversity quality of recommendations than the diversity measure in automatic evaluation. The Intent-Coverage is formally defined as

$$\text{Intent-Coverage}(k) = \frac{1}{I}\sum_{i}^{I} B_i(k), \tag{9}$$

where $B_i(k)$ is a binary variable indicating whether the intent i is found within the top k recommendations or not, and I is the total number of unique intents for each test query.

4.3.4 Qualitative Comparison

Similarly, we first show a couple of qualitative comparisons to gain some intuition on the difference between the queries recommended by our approach and other baselines. Table 6 shows two samples from our test queries including their top 10 recommendations generated by five methods.

Except for the baselines described in Section 4.1, we also adopt Baseline-Naive for query recommendation. Baseline-Naive only measures relevance of query recommendations using Euclidean distance between queries, without considering diversity.

From the results in Table 6, we clearly see that the Baseline-Naive approach tends to recommend highly related but somewhat redundant queries. For example, for the test query "abc," we can find equivalent recommendations like "abc tv" and "abc television," or recommendations sharing very close meaning like "abc news," "abc breaking

[7] TREC Web Track: http://plg.uwaterloo.ca/ trecweb.

TABLE 7
Performance of Recommendation Results over a Sample of Queries under Five Different Approaches

	α-nDCG@5	α-nDCG@10	Intent-Coverage@5	Intent-Coverage@10
Baseline-Naive	0.717	0.689	0.300	0.536
Baseline-MR	0.758 (5.7‡/*/*/*)	0.703 (2.0/*/*/*)	0.353 (17.7‡/*/*/*)	0.541 (0.9/*/*/*)
Baseline-MMR	0.799 (11.4‡/5.4‡/*/*)	0.742 (7.7‡/5.5‡/*/*)	0.384 (28‡/8.7/*/*)	0.585 (9.1‡/8.1‡/*/*)
Baseline-GH	0.794 (10.7‡/4.7‡/-0.6/*)	0.768 (11.5‡/9.2‡/3.5‡/*)	0.373 (24.3‡/5.6/-2.9/*)	0.616 (14.9‡/13.8‡/5.3/*)
MRSP	0.838 (16.9‡/10.5‡/4.9‡/5.5‡)	0.806 (17‡/14.6‡/8.6‡/4.9‡)	0.436 (45.3‡/23.5‡/13.5‡/16.9‡)	0.665 (24.1‡/22.9‡/13.7‡/8‡)

Performance metrics α-nDCG@5, α-nDCG@10, Intent-Coverage@5, and Intent-Coverage@10 are shown. Numbers in parentheses indicates relative percent improvement over Baseline-Naive/Baseline-MR/Baseline-MMR/Baseline-GH. Paired t-tests are performed, and results which show significant improvements (p-value < 0.05) are marked ‡.

news," and "abc world news." We can also find redundant examples in the recommendations for the test query "yamaha," e.g., "yamaha motor," "yamaha motorcycle," and "yamaha motorcycles." Since Baseline-Naive method only considers relevance, it will inevitably produce many redundant recommendations. On the other hand, if we further go down the recommendation lists provided by Baseline-Naive or Baseline-MMR, we will notice that they may bring up long-tail queries which are not so important or representative for recommendation, e.g., recommendation "scooter trade yamaha north america" for test query "yamaha" at rank 14 under Baseline-Naive and 19 under Baseline-MMR, respectively.

Meanwhile, we can easily find that the three baseline approaches recommend queries with better diversity than the naive method. However, there is still some redundancy in these approaches. For example, for the test query "abc," "abc tv" and "abc television" are both recommended by Baseline-MMR, while for test query "yamaha," "yamaha motor," and "yamaha motorcycles" are both recommended by Baseline-MR and Baseline-GH.

Among all these approaches, we observe that our MRSP approach obtains best performance, where more diverse, relevant and representative queries can be found in its recommendation results.

4.3.5 Quantitative Comparison

In our experiments, we compare the performance of different methods in terms of α-nDCG@5, α-nDCG@10, Intent-Coverage@5, and Intent-Coverage@10. Table 7 reports the performance of different recommendation approaches under manual evaluation. All the metrics here take 1 as its upper bound (i.e., the best case), and 0 as its lower bound (i.e., the worst case). The numbers in the parentheses are the relative improvements compared with baseline methods.

From Table 7, we can see that Baseline-Naive obtains the lowest Intent-Coverage and also shows a poor overall performance as measured by α-nDCG, since it only consider the relevance in recommendation. Baseline-MR gets better performance on Intent-Coverage and α-nDCG than Baseline-Naive. The major reason is that Baseline-MR tends to assign relevant and representative queries higher ranking scores. With additional diversity penalty steps, Baseline-MR also achieves diversity. Both Baseline-MMR and Baseline-GH can further outperform Baseline-Naive on α-nDCG by explicitly addressing recommendation diversity. Baseline-GH can achieve better performance than Baseline-MMR when the recommendation size is large. Compared with the four baseline methods, our MRSP approach achieves the best performance in terms of all measures, which is consistent with results reported in the automatic evaluation. We also conduct a t-test (p-value $<$ 0.05) and find that the improvements over all baseline methods are significant. It shows that by exploiting the intrinsic global query manifold structure and employing manifold ranking with sink points, we can recommend highly diverse as well as highly related queries.

5 Conclusion

In this paper, we propose a novel MRSP approach to address diversity as well as relevance and importance in ranking. MRSP uses a manifold ranking process over the data manifold, which can naturally find the most relevant and important objects. Meanwhile, by turning ranked objects into sink points on data manifold, 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 competitions and outperform other baseline methods. Experiments for query recommendation also demonstrate that our approach can effectively generate diverse and highly relevant query recommendations.

Acknowledgments

This research work was funded by the National Grand Fundamental Research 973 Program of China under Grant No. $2012CB316303$, the National High-tech R&D Program of China under Grant No. $2010AA012500$, and the National Natural Science Foundation of China under Grant Nos. 61003166 and 60933005.

References

[1] R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong, "Diversifying Search Results," *Proc. Second ACM Int'l Conf. Web Search and Data Mining (WSDM '09),* pp. 5-14, 2009.

[2] J. Allan, R. Gupta, and V. Khandelwal, "Temporal Summaries of News Topics," *Proc. 24th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '01),* pp. 10-18, 2001.

[3] D. Beeferman and A. Berger, "Agglomerative Clustering of a Search Engine Query Log," *Proc. Sixth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining,* pp. 407-416, 2000.

[4] P. Boldi, F. Bonchi, C. Castillo, D. Donato, and S. Vigna, "Query Suggestions Using Query-Flow Graphs," *Proc. Workshop Web Search Click Data (WSCD '09),* pp. 56-63, 2009.

[5] F. Boudin, M. El-Bèze, and J.-M. Torres-Moreno, "A Scalable MMR Approach to Sentence Scoring for Multi-Document Update Summarization," *Proc. Companion Vol.: Posters (Coling '08),* pp. 23-26, Aug. 2008.

[6] J. Carbonell and J. Goldstein, "The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries," *Proc. 21st Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '98),* pp. 335-336, 1998.

[7] C.L. Clarke, M. Kolla, G.V. Cormack, O. Vechtomova, A. Ashkan, S. Büttcher, and I. MacKinnon, "Novelty and Diversity in Information Retrieval Evaluation," *Proc. 31st Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval,* pp. 659-666, 2008.

[8] H.T. Dang and K. Owczarzak, "Overview of the TAC 2009 Summarization Track (Draft)," *Proc. Second Text Analysis Conf. (TAC '09),* 2009.

[9] P.G. Doyle and J.L. Snell, *Random Walks and Electric Networks,* Math. Assoc. of Am., 1984.

[10] P. Du, J. Guo, J. Zhang, and X. Cheng, "Manifold Ranking with Sink Points for Update Summarization," *Proc. 19th ACM Conf. Information and Knowledge Management (CIKM '10),* 2010.

[11] G. Erkan and D.R. Radev, "Lexrank: Graph-Based Lexical Centrality as Salience in Text Summarization," *J. Artificial Int'l Research,* vol. 22, no. 1, pp. 457-479, 2004.

[12] T.H. Haveliwala, "Topic-Sensitive Pagerank," *Proc. 11th Int'l Conf. World Wide Web,* pp. 517-526, 2002.

[13] K. Järvelin and J. Kekäläinen, "Cumulated Gain-Based Evaluation of IR Techniques," *ACM Trans. Information Systems,* vol. 20, no. 4, pp. 422-446, 2002.

[14] K. Knight and D. Marcu, "Statistics-Based Summarization - Step One: Sentence Compression," *Proc. 17th Nat'l Conf. Artificial Intelligence and 12th Conf. Innovative Applications of Artificial Intelligence,* pp. 703-710, 2000.

[15] Y. Lan, T.-Y. Liu, Z. Ma, and H. Li, "Generalization Analysis of Listwise Learning-to-Rank Algorithms," *Proc. 26th Ann. Int'l Conf. Machine Learning (ICML '09),* pp. 577-584, 2009.

[16] L. Li, Z. Yang, L. Liu, and M. Kitsuregawa, "Query-URL Bipartite Based Approach to Personalized Query Recommendation," *Proc. 23rd Nat'l Conf. Artificial Intelligence,* pp. 1189-1194, 2008.

[17] W. Li, F. Wei, Q. Lu, and Y. He, "PNR2: Ranking Sentences with Positive and Negative Reinforcement for Query-Oriented Update Summarization," *Proc. 22nd Int'l Conf. Computational Linguistics (Coling '08),* pp. 489-496, Aug. 2008.

[18] C.-Y. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries," *Proc. ACL Workshop Text Summarization Branches Out,* pp. 74-81, July 2004.

[19] C.-Y. Lin and E. Hovy, "Manual and Automatic Evaluation of Summaries," *Proc. ACL-02 Workshop Automatic Summarization,* pp. 45-51, 2002.

[20] Q. Mei, J. Guo, and D. Radev, "DivRank: The Interplay of Prestige and Diversity in Information Networks," *Proc. 16th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '10),* pp. 1009-1018, 2010.

[21] Q. Mei, D. Zhou, and K. Church, "Query Suggestion Using Hitting Time," *Proc. 17th ACM Conf. Information and Knowledge Management,* pp. 469-477, 2008.

[22] R. Mihalcea and P. Tarau, "TextRank: Bringing Order into Texts," *Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP '04),* pp. 404-411, July 2004.

[23] J. Otterbacher, G. Erkan, and D. Radev, "Using Random Walks for Question-Focused Sentence Retrieval," *Proc. Human Language Technology Conf. and Conf. Empirical Methods in Natural Language Processing,* pp. 915-922, Oct. 2005.

[24] D.R. Radev, H. Jing, and M. Budzikowska, "Centroid-Based Summarization of Multiple Documents: Sentence Extraction, Utility-Based Evaluation, and User Studies," *Proc. Workshop Automatic Summarization (NAACL-ANLP '00),* pp. 21-30, 2000.

[25] D.R. Radev and K.R. McKeown, "Generating Natural Language Summaries from Multiple On-Line Sources," *Computational Linguistics,* vol. 24, pp. 470-500, Sept. 1998.

[26] S. Roweis and L. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science,* vol. 290, pp. 2323-2326, 2000.

[27] R.L. Santos, C. Macdonald, and I. Ounis, "Exploiting Query Reformulations for Web Search Result Diversification," *Proc. 19th Int'l Conf. World Wide Web (WWW '10),* pp. 881-890, 2010.

[28] J.B. Tenenbaum, V. de Silva, and J.C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science,* vol. 290, pp. 2319-2323, 2000.

[29] X. Wan, "TimedTextRank: Adding the Temporal Dimension to Multi-Document Summarization," *Proc. 30th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '07),* pp. 867-868, 2007.

[30] X. Wan, J. Yang, and J. Xiao, "Manifold-Ranking Based Topic-Focused Multi-Document Summarization," *Proc. 20th Int'l Joint Conf. Artificial Intelligence (IJCAI '07),* pp. 2903-2908, Jan. 2007.

[31] F. Wei, W. Li, Q. Lu, and Y. He, "Query-Sensitive Mutual Reinforcement Chain and Its Application in Query-Oriented Multi-Document Summarization," *Proc. 31st Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '08),* pp. 283-290, 2008.

[32] J.-R. Wen, J.-Y. Nie, and H.-J. Zhang, "Clustering User Queries of a Search Engine," *Proc. 10th Int'l Conf. World Wide Web,* pp. 162-168, 2001.

[33] C.X. Zhai, W.W. Cohen, and J. Lafferty, "Beyond Independent Relevance: Methods and Evaluation Metrics for Subtopic Retrieval," *Proc. 26th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '03),* pp. 10-17, 2003.

[34] B. Zhang, H. Li, Y. Liu, L. Ji, W. Xi, W. Fan, Z. Chen, and W.-Y. Ma, "Improving Web Search Results Using Affinity Graph," *Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '05),* pp. 504-511, 2005.

[35] J. Zhang, X. Cheng, G. Wu, and H. Xu, "AdaSum: An Adaptive Model for Summarization," *Proc. 17th ACM Conf. Information and Knowledge Management (CIKM '08),* pp. 901-910, 2008.

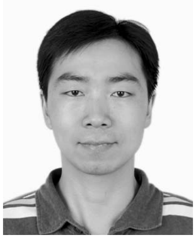
[36] D. Zhou, O. Bousquet, T.N. Lal, J. Weston, and B. Schölkopf, "Learning with Local and Global Consistency," *Advances in Neural Information Processing Systems 16,* S. Thrun, L. Saul, and B. Schölkopf, eds., MIT Press, 2004.

[37] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf, "Ranking on Data Manifolds," *Advances in Neural Information Processing Systems 16.* S. Thrun, L. Saul, and B. Schölkopf, eds., MIT Press, 2004.

[38] X. Zhu, A. Goldberg, J. Van Gael, and D. Andrzejewski, "Improving Diversity in Ranking Using Absorbing Random Walks," *Proc. NAACL-HLT,* pp. 97-104, Apr. 2007.

[39] X. Zhu, J. Guo, and X. Cheng, "Recommending Diverse and Relevant Queries with a Manifold Ranking Based Approach," *Proc. Workshop Query Representation and Understanding (SIGIR '10),* 2010.

Xue-Qi Cheng is a professor at the Institute of Computing Technology, Chinese Academy of Sciences (ICT-CAS), and the director of the Research Center of Web Data Science and Engineering (WDSE) in ICT-CAS. His main research interests include network science, web search and data mining, P2P, and distributed system, information security. He has published more than 100 publications in prestigious journals and international conferences, including New *Journal of Physics*, *Journal of Statistics Mechanics: Theory and Experiment*, *IEEE Transactions on Information Theory*, ACM SIGIR, WWW, ACM CIKM, WSDM, AAAI IJCAI and so on. He is currently serving on the editorial boards of the *Journal of Computer Science and Technology*, *Journal of Computer Research and Development*, and *Journal of Computer*. He is a recipient of the Youth Science and Technology Award of Mao Yisheng Science and Technology Award (2008), the CVIC Software Engineering Award (2008), the Excellent Teachers Award from Graduate University of Chinese Academy of Sciences (2006), the Second prize for the National Science and Technology Progress (2004), the Second prize for the Chinese Academy of Sciences' Science and Technology Progress (2002), and the Young Scholar Fellowship of Chinese Academy Sciences (2000). He is a member of the IEEE.



Pan Du received the BS degree in environmental engineering from Yanshan University in 2003 and the MS degree in computer science from Beijing Institute of Technology in 2006. Currently, he is working toward the PhD degree at the Institute of Computing Technology, Chinese Academy of Sciences, China. His research interests include web search and mining, text analysis, and machine learning. He is a student member of the IEEE.

Jiafeng Guo received the BE degree in computer science and technology from the University of Science and Technology of China, Hefei, in 2004, and the PhD degree in computer software and theory from the Institute of Computing Technology, the Chinese Academy of Sciences (CAS), Beijing, China, in 2009. Currently, he is working as an assistant researcher in the Institute of Computing Technology, Chinese Academy of Sciences. His major research interests include Web search and mining, user data mining, machine learning, and social network. He has published more than 20 papers in international conferences and journals, including the top conferences in IR like SIGIR, WWW, and CIKM. He currently serves on the program committees of several international conferences, including SIGIR, EMNLP, and AIRS. He was the recipient of the Excellence PhD Dissertation Award of CAS in 2009.

Xiaofei Zhu received the BS and MS degrees in computer science from the Chongqing University of Technology, China, in 2003, and Chongqing University of Posts and telecommunications, China, in 2008, respectively. Currently, he is working toward the PhD degree in the Institute of Computing Technology at Chinese Academy of Science. His research interests include Web search, query understanding, and data mining.

Yixin Chen received the PhD degree in computing science from the University of Illinois at Urbana-Champaign in 2005. Currently, he is working as an associate professor of computer science at the Washington University in St. Louis, Missouri. His research interests include nonlinear optimization, constrained search, planning and scheduling, data mining, and data warehousing. His work on planning has won First Prizes in the International Planning Competitions (2004 and 2006). He has won the Best Paper Award at AAAI (2010) and ICTAI (2005), and Best Paper nomination at KDD (2009). He has received an Early Career Principal Investigator Award from the Department of Energy (2006) and a Microsoft Research New Faculty Fellowship (2007). He serves on the Editorial Board of the *IEEE Transactions of Knowledge and Data Engineering* and *Journal of Artificial Intelligence Research*. He is a senior member of the IEEE.