*Applied Mechanics and Materials Vols 380-384 (2013) pp 2811-2816* *Online: 2013-08-30*
*© (2013) Trans Tech Publications, Switzerland*
*doi:10.4028/www.scientific.net/AMM.380-384.2811*

# A Novel Biased Diversity Ranking Model for Query-Oriented Multi-document Summarization

Kai Lei*, Yifan Zeng

The Shenzhen Key Lab for Cloud Computing Technology and Application

Peking University Shenzhen Graduate School, Shenzhen, Guangdong 518055, P.R. China

Email: leik@pkusz.edu.cn*, zengyifang@sz.pku.edu.cn

**Keywords:** multi-document summarization; manifold-rankingprocess; divrank

**Abstract.** Query-oriented multi-document summarization (QMDS) attempts to generate a concise piece of text byextracting sentences from a target document collection, withthe aim of not only conveying the key content of that corpus,also, satisfying the information needs expressed by that query.Due to its great applicable value, QMDS has been intensively studied in recent decades. Three properties are supposed crucial for a good summary, i.e., relevance, prestige and low redundancy (orso-called diversity). Unfortunately, most existing work either disregarded the concern of diversity, or handled it with non-optimized heuristics, usually based on greedy sentences election. Inspired by the manifold-ranking process, which deals with query-biased prestige, and DivRank algorithm which captures query-independent diversity ranking, in this paper, we propose a novel biased diversity ranking model, namedManifoldDivRank, for query-sensitive summarization tasks. The top-ranked sentences discovered by our algorithm not only enjoy query-oriented high prestige, more importantly, they are dissimilar with each other. Experimental results on DUC2005and DUC2006 benchmark data sets demonstrate the effectivenessof our proposal.

## INTRODUCTION

With the explosive growth of the internet, users usually get lost in the overwhelming information available online. Therefore, it will save much time of users if certain retrieval system could present the major content of source information. Query-oriented multi-document summarization (QMDS) condenses large amount of target documents, for the purpose of using the resulted text to meet the user's information needs. Hence, QMDS has become one of the hottest researches in Natural Language Processing and Information Retrieval Communities.Such a task is commonly treated as a sentence ranking problem.

Using a short piece of text instead of the whole target document collection to answer a user's complex question (i.e., query), three attributes are recognized to be crucial judgments of a good summary: relevance, prestige and diversity. "Relevance"means that selected sentences must be relevant with the query; "prestige" indicates that those sentences should be prestigious among all the candidate sentences; and "diversity" suggests that summary sentences should have low degree of overlapping content. Whereas, most pioneering works either disregarded partial above properties or captured them with an un-unified selection process. For example, Although Shen et al. [1] was able to develop ranking SVM, a pair-wise learning to rank framework [2], to learn the ranking scores of sentences in QMDS task, they had to use an extra diversity penalty strategy presented in [3] to control redundancy in subsequent sentence selection process. Similarly, Li et al. [4] succeeded in designing a linear combination to incorporate the query-biased relevance, novelty and coverage. However, those sentence properties were still derived in miscellaneous greedy processes. Hence, how to form a unified ranking algorithm which enables us to dig out query-aware relevant, prestigious and mutually dissimilar sentences gives us strong motivation for further study.

The work in [3] is the first attempt to employ manifold-rankingprocess [5] for QMDS tasks. The rationale of manifold-ranking process is as follows: firstly construct agraph where sentences and the query description denote vertices/nodes, and sentence similarity is embodied as edge weights; then the query node is endowed initial score 1 while other nodes are given 0; based on such initialization, each graph node iteratively propagates its score to all neighbors along those weighted edges until all the node scores keep unchanged. Manifold-ranking process can easily find the most relevant and prestigious sentences towards the query. Whereas, it fails to make the top-ranked sentences dissimilar with each other. Differently, Mei et al. [6] came up with DivRank, by developing a vertex-reinforced random walk over an adjacent graph, to conduct a comprehensive quantification of objects with regard to their prestige as well as diversities. Unfortunately, DivRank is query-independent and is not applicable to QMDS tasks where query has to be taken into consideration. Inspired by above two researchers, in this work, we propose a novel query-biased diversity ranking model, named ManifoldDivRank, for QMDS through combining the principles of both manifold-ranking process and DivRank approach.

To be specific, our ManifoldDivRank is a graph-based ranking algorithm. Similar with the settings of the traditional manifold-ranking process, we also set 1 and 0 as the initial scores of query and other sentences, respectively. Then, during the iteration, we use the real-time accumulative visit times of each node to update the incoming information amount from its neighbors, which, likes the principles of DivRank, enables us to enlarge the score gaps of near sentences. After the iterative process converges, we could directly select some top-ranked sentences to construct a summary without any extra operations. Experimental results on standard summarization data sets DUC2005-2006 demonstrate the good performance of our ManifoldDivRank in improving the quality of summaries.

The rest of this paper is organized as follows: Section 2 introduces related work. Section 3 gives the detail of sentence ranking using our ManifoldDivRank Model. Section 4 shows experiments and results. Finally, we conclude our work in Section 5.

## Related work

As our work focuses on query-oriented multi-document summarization, next, we mainly introduce some typical literature in this area.

The most representative QMDS approaches depend on graph-based ranking algorithms, where the similar principles of PageRank [7] and HITS [8] have been successfully applied. Topic-sensitive LexRank [9] has been applied to the task of query-focused summarization where the relevance of asentence to the query is taken into account when performing a dom walk. Mihalcea et al. [10] independently proposed another graph-based random walk model TextRank, which is similar to LexRank, except that it is applied to single-document summarization, and it does not consider similarity threshold in sentence extraction. Wan et al. [3] applied a manifold-ranking algorithm to query-focused summarization which can simultaneously make full use of both the relationships among all the sentences in the documents and the relationships between the given query and the sentences. Furthermore, some works concentrated on the feature design. Here, “feature” mainly indicates the desirable propertiesof a good summary, such as the relevance, prestige/centrality, diversity/novelty, and etc. Those authors wanted to directly select some top-ranked sentences through designing skillful sentence features. Take as an example from the work in [4], researchers used various greedy strategies to design some query-aware sentence features, and finally scored and ranked candidate sentences over a combination of those component feature scores. Look at again the study in [11], Yin et al. provided six features, including four kinds of relevance, biased information richness and biased information novelty, and finally developeda Gaussian Ranking Model to sort all sentences.

Moreover, supervised learning approaches have been successfully applied to query/topic-biased summarization. Ouyang et al. [12] used support vector regression (SVR), a point-wise ranking algorithm, to relate the “true” score of the sentence to its features. Jin et al. [13] presented a systematic study of comparing different learning to rank algorithms and comparing different selection strategies for multi-document summarization. However, it focused on the simple comparison of some basic models with no optimization. Work [14] had deeply investigated and compared the effects of using different automatic annotation techniques on different supervised learning approaches, including SVMs, HMMs, CRFs, and MaxEnt, in the domain of query-focused multi-document summarization. Research [1] explored the use of ranking SVM, a pair-wise learning to rank model, for obtaining credible and controllable solutions for feature combinations.

## SENTENCE RANKING USING MANIFOLDDIVRANK

Let points $\{x_0, x_1, \ldots, x_n\}$ denote the query statement $(x_0)$ and all the sentences in the document collection $(\{x_i|1 \leq i \leq n\})$ in a manifold space. Construct a connectivity graph $G = (N, S)$ where $N = (x_i|0 \leq i \leq n)$ is the node set, and $S$ is a $n \times n$ symmetric similarity matrix ($S_{i,j}$ stands for the similarity score between sentence $x_i$ and sentence $x_j$). Since $x_0$ denotes the query description, the initial score vector of these sentences is defined as $Prob^*$, where $Prob^*(x_0) = 1$ and $Prob^*(x_i) = 0\ (1 \leq i \leq n)$. Accordingly, the transition probability from node $x_i$ to $x_j$ in our ManifoldDivRank iterative process can be represented using following formula:

$$\mathrm{Prob}_t(x_j|x_i) = (1-\alpha)\cdot \mathrm{Prob}^*(x_j) + \alpha \cdot \frac{\mathrm{Prob}_0(x_j|x_i)\cdot N_t(x_j)}{D_t(x_i)} \quad (1)$$

Where

$$D_t(x_i) = \sum_{x_j \in N \backslash x_i} \mathrm{Prob}_0(x_j|x_i)\, N_t(x_j) \quad (2)$$

Here, parameter $\alpha$ is used to balance the impacts of prior knowledge (i.e., $Prob^*$) and the information spreads through the graph structure. Its influence will be discussed in experimental part. In addition, $N_t(x_j)$ counts the visiting times of node $x_j$ in the vertex-reinforced random walk up to time *t*.

$Prob_0(x_j|x_i)$ is the "organic" transition probability priorto any reinforcement, which is derived from the primitive network structure:

$$\mathrm{Prob}_0(x_j|x_i) = \frac{\bar{S}_{i,j}}{\sum_j \bar{S}_{i,j}} \quad (3)$$

Here, we assume that matrix *S* is irreducible, namely, each pair of sentences could be connected via a route. It is a reasonable assumption and was also adopted by literature [7] in the PageRank matrix for calculating PageRank values.

If the sentence graph is ergodic, after a sufficient large time *T*, the iterative process defined in Equation (1) will converges to a stationary distribution p. That is

$$p(x_j) = \sum_{x_t \in N \backslash x_j} \mathrm{Prob}_t(x_j|x_i)\, p(x_i)\ \forall t \geq T \quad (4)$$

$p(.)$ is then used as the final ranking scores of sentences, which is a balancing qualification of the query-biased relevance, prestige and diversity. Apparently, $\sum_i p(x_i) = 1$.

### *A. Efficient Approximation*

In above subsection, we give the general form of ManifoldDivRank based on a general random walk. Note that the expectation of $N_t(x_i)$ follows the recurrent formula:

$$E[N_{t+1}(x_i)] = E[N_t(x_i)] + \mathrm{Prob}_{t+1}(x_i) \quad (5)$$

However, in ManifoldDivRank, $Prob_{t+1}(x_i)$ relies on $N_t(x_i)$ and tracking $N_t(x_i)$ is non-trivial. Efficient approximation is needed for practical application. Just like the work in [6], we also use accumulative incoming probabilities to replace the sum of visiting times, i.e.

$$N_T(x_i) \propto \sum_{t=0}^{T} \mathrm{Prob}_t(x_i) \quad (6)$$

### *B. Analytical Discussion*

Above ManifoldDivRank formula integrates the principles of both manifold-ranking process and DivRank. After the settings of initial scores, graph nodes spread their influence to their near neighbors. Manifold-ranking algorithm makes use of such method to find query-aware prestigious sentences while failing to penalize the acquired scores of related sentences. Our modification, using vertex-reinforced random walk, "screws" the primitive network structure. At time 0, Equation (1) favors nodes with a higher centrality. As time goes, the nodes which already have a high $N_t(.)$ tend to get an even higher weight, i.e., there emerges a **rich-gets-richer** phenomenon. In other words, the initialization and information propagation on thesentence connectivity graph are able to guarantee the biased relevance and high prestige while vertex-reinforced randomwalk aims to enhance the score difference of close sentences. Therefore, our model could skillfully finds some desired sentences which exhibit query-sensitive relevance, prestige and diversity simultaneously.

## EXPERIMENTS

*A. Data sets and evaluation metrics*

We conduct experiments on the main tasks of benchmark summarization datasets DUC2005[1] and DUC2006[2]. Each task has a gold standard data set and reference summaries. Table I gives a short summary ofabove data sets. Documents are pre-processed by segmenting sentences and splitting words. Stop words are removed and the remaining words are stemmed using Porter stemmer[3].

We use the ROUGE [15] (version 1.5.5) toolkit[4]for evaluation, which is officially adopted by Document Understanding Conference (DUC) for evaluating automatically generated summaries. It includes five methods: ROUGE-N, ROUGE-L,ROUGE-W, ROUGE-S and ROUGE-SU. ROUGE-N uses *n*-grams recall between the candidate summary and the reference summary, and is computed as follows:

$$ROUGE - N = \frac{\sum_{s \in S_R} \sum_{n \in s} N_m(n)}{\sum_{s \in S_R} \sum_{n \in s} N(n)} \quad (7)$$

TABLE I. SUMMARY OF DATA SETS

| | **DUC2005** | **DUC2006** |
|---|---|---|
| Task | the only task | the only task |
| Number of documents | 1593 | 1250 |
| Number of clusters | 50 | 50 |
| Data source | TREC | TREC |
| Summary length | 250 words | 250 words |

Where $n$ denotes $n$-grams and $N_m(n)$ is the maximum number of $n$-grams co-occurring in a candidate summary and a set of reference summaries. The sentence set$S_R$ is the reference summaries and $s$ is a sentence in the set.$N(n)$is the number of $n$-grams in the set of reference summaries. The length of longest common subsequence (LCS) between candidate summaries and reference summaries is employed by ROUGE-L and ROUGE-W uses the weighted LCS. Here were port the average *F*-measure scores of ROUGE-1, ROUGE-2 and ROUGE-SU4, which base on Uni-gram match, Bi-gram match, and unigram plus skip-bigram match with maximumskip distance of 4 between the candidate summary and the reference summary, respectively.

*B. Experimental Results and Analysis*

To validate our ManifoldDivRank in query-biased task, following systems are implemented as baselines. (1)**Random**: the method selects sentences randomly for each document collection. (2)**Manifold** [3]: ranking the sentences according to the manifold ranking scores and uses the greedy algorithm similar to [16] to impose the diversity penalty. (3)**BiasedLexRank**: a modified version of traditional random walk with prior belief. It was presented in [17] and the sentence prior knowledge was denoted by sentence's similarity tothe query description. (4)**DivRank** [6]: we incorporate the basic DivRank model with query-sentence similarities as those sentences' prior distribution, which biases the final ranking list towards the query. Tables II and III present the performance of these systems on DUC2005 and DUC2006 data sets, respectively.

TABLE II. *F*-MEASURE COMPARISON ON DUC2005

| **Systems** | **ROUGE-1** | **ROUGE-2** | **ROUGE-4** |
|---|---|---|---|
| Random | 0.30821 | 0.03976 | 0.10625 |
| Biased LexRank | 0.36993 | 0.07113 | 0.12805 |
| DivRank | 0.37324 | 0.07286 | 0.12937 |
| Manifold | 0.37497 | 0.07423 | 0.12907 |
| ManifoldDivRank | **0.38193** | **0.07644** | **0.13297** |

[1]http://www-nlpir.nist.gov/projects/duc/duc2005/tasks.html
[2]http://www-nlpir.nist.gov/projects/duc/duc2006/tasks.html
[3]http://tartarus.org/ martin/PorterStemmer/
[4]http://www.isi.edu/licensed-sw/see/rouge/

Apparently, our approach ManifoldDivRank has the optimal performance, especially outperforming the three representative graph-based methods: *Biased LexRank*, *DivRank* and *Manifold*. Note that both Biased LexRank and Manifold aim to rank candidate sentences based on their relevance and prestige/centralities, and resort to greedy strategy, Maximum Marginal Relevance (MMR) algorithm, to control redundancy during sentence selection. Such kinds of methods are difficult to balance the prestige and diversity well, and their performances are easily to be affected by those greedy selection algorithms.

TABLE III. *F*-MEASURE COMPARISON ON DUC2006

| Systems | ROUGE-1 | ROUGE-2 | ROUGE-4 |
|---|---|---|---|
| Random | 0.34821 | 0.05297 | 0.11908 |
| Biased LexRank | 0.36814 | 0.08074 | 0.12993 |
| DivRank | 0.37973 | 0.08785 | 0.13162 |
| Manifold | 0.38867 | 0.08308 | 0.13307 |
| ManifoldDivRank | **0.40852** | **0.09348** | **0.15631** |

DivRank approach is more applicable to query-independent tasks [6]. Even though our work attempted to add query-oriented similarities as sentence prior distribution, DivRank constantly screws the original network structure, resulting that it focuses on diversity too much, at the cost of sacrificing the prestige of the entire ranking list to some extent.

*C. Influence of Parameter$\alpha$*

In Equation (1), we set parameter $\alpha$ to control the impacts of prior distribution and information coming from neighbors. Obviously, when $\alpha = 0$, it means that we only considered the prior knowledge and took no account of inter-relationships among sentences. Contrarily, $\alpha = 1$ suggests that we neglect the influence of prior knowledge thoroughly.

In order to investigate the influence of parameter, we vary $\alpha$ from 0 to 1 with an interval of 0.5. Figure 1 demonstrates the trend of summary quality (measured by ROUGE-1) along with $\alpha$ over both DUC2005 and DUC2006 data sets.

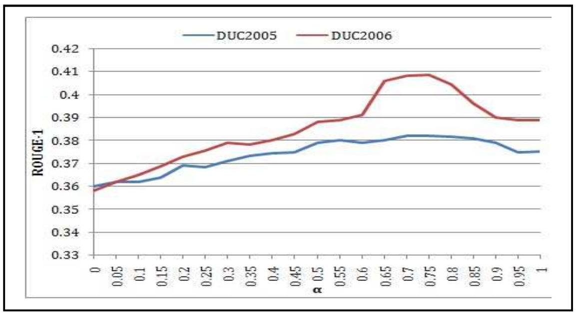


Figure 1. ROUGE-1 vs. Parameter$\alpha$

From above figure, we could clearly find that parameter $\alpha$ indeed influences the performance of our ManifoldDivRank model in summarization tasks. Overall, the model performs poorest when $\alpha = 0$, which means that our consideration of interactions among sentences benefits the discovering of truly prestigious sentences as well as redundancy reduction. With the increase of $\alpha$, evaluation scores rise and reach respective maximum around $\alpha = 0.75$, then drop gradually when $\alpha$ continues to become larger.

## CONCLUSIONS AND FUTURE WORK

In this work, we proposed a novel diversity ranking model, ManifoldDivRank, for query-oriented multi-document summarization task. Our model is inspired by two basic algorithms: manifold-ranking process and DivRank algorithm. The former is able to capture query-sensitive relevant and prestigious sentences while failing to integrate diversity property. Contrarily, the latter concentrates on diversity while cannot be applied in query-dependent scenarios. Our proposed ManifoldDivRankmodel enjoys their strengths, hence, enables us to find biased relevant, prestigious and diverse sentences for constructing high-quality summaries.

**Acknowledgments**

This research is financially supported by NSFC (61103027), 973 Project (2011CB302305), Guangdong Gov Projects (2011A090200063 and 2011B090400396), and Shenzhen Gov Projects (JCYJ20120829170028558 and ZYA201106080025A）

**References**

[1] C.Shen and T.Li "Learning to rank for query-focused multi-document summarization," in Data Mining (ICDM), 2011 IEEE 11th InternationalConference on. IEEE, 2011, pp. 626–634.

[2] Z.Cao, T.Qin, T.Liu, M.Tsai, and H.Li, "Learning to rank: frompairwise approach to listwise approach," in Proceedings of the 24th International Conference on Machine Learning. ACM, 2007, pp. 129–136.

[3] X. Wan, J. Yang, and J. Xiao, "Manifold-ranking based topic-focusedmulti-document summarization," in Proceedings of the 20th International Joint Conference on Artifical Intelligence. Morgan KaufmannPublishers Inc., 2007, pp. 2903–2908.

[4] X. Li, Y. Shen, L. Du, and C. Xiong, "Exploiting novelty, coverageand balance for topic-focused multi-document summarization," in Proceedings of the 19th ACM International Conference onInformation and Knowledge Management. ACM, 2010, pp. 1765–1768.

[5] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf,"Ranking on data manifolds," Advances in Neural InformationProcessing Systems, vol. 16, pp. 169–176, 2004.

[6] Q. Mei, J. Guo, and D. Radev, "Divrank: the interplay of prestige anddiversity in information networks," in Proceedings of the 16th ACMSIGKDD International Conference on Knowledge Discovery andData Mining. ACM, 2010, pp. 1009–1018.

[7] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citationranking: bringing order to the web." 1999.

[8] J. Kleinberg, "Authoritative sources in a hyperlinked environment," Journal of the ACM (JACM), vol. 46, no. 5, pp. 604–632, 1999.

[9] J. Otterbacher, G. Erkan, and D. Radev, "Using random walks forquestion-focused sentence retrieval," in Proceedings of theConferenceon Human Language Technology and Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2005,pp. 915–922.

[10] R. Mihalcea and P. Tarau, "Textrank: Bringing order into texts," inProceedings of EMNLP, vol. 4. Barcelona, Spain, 2004, pp. 404–411.

[11] W. Yin, Y. Pei, F. Zhang, and L. Huang, "Query-focused multi-documentsummarization based on query-sensitive feature space," in Proceedingsof the 21st ACM International Conference on Information and KnowledgeManagement. ACM, 2012, pp. 1652–1656.

[12] Y. Ouyang, S. Li, and W. Li, "Developing learning strategies fortopic-basedsummarization," in Proceedings of the sixteenth ACMConference on Conference on Information and KnowledgeManagement. ACM,2007, pp. 79–86.

[13] F. Jin, M. Huang, and X. Zhu, "A comparative study on ranking andselection strategies for multi-document summarization," in Proceedingsof the 23rd International Conference on ComputationalLinguistics:Posters. Association for Computational Linguistics, 2010, pp. 525–533.

[14] Y. Chali and S. Hasan, "Query-focused multi-document summarization:automatic data annotations and supervised learningapproaches," NaturalLanguage Engineering, vol. 18, no. 1, p. 109, 2012.

[15] C. Lin, "Rouge: A package for automatic evaluation of summaries,"in Text Summarization Branches Out: Proceedings of the ACL-04Workshop, 2004, pp. 74–81.

[16] B. Zhang, H. Li, Y. Liu, L. Ji, W. Xi, W. Fan, Z. Chen, and W. Ma,"Improving web search results using affinity graph," in Proceedings ofthe 28th AnnualInternational ACM SIGIR Conference on Research andDevelopment in Information Retrieval. ACM, 2005, pp. 504–511.

**Vehicle, Mechatronics and Information Technologies**
10.4028/www.scientific.net/AMM.380-384

**A Novel Biased Diversity Ranking Model for Query-Oriented Multi-Document Summarization**
10.4028/www.scientific.net/AMM.380-384.2811