www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume - 3 Issue - 8 August, 2014 Page No. 7834-7836

Ranking on Data Manifold with Sink Points

R.Kanimozhi, Dr. Y. Kalpana

Associate Professor
Vels University, Pallavaram - 117.

Abstract: Ranking is an important problem in various applications. Many natural language processing tasks involve ranking a set of items. Sometimes we want the top items to be not only good individually but also *diverse* collectively. These ranking approaches are used to avoid redundant information as possible. Manifold Ranking with Sink Points (MRSP), is used to address the diversity and importance in ranking. We applied MRSP on two application tasks, update summarization and query recommendation. This ranking approach gives a strong performance of MRSP as compared to already 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. A mass of relevant objects may contain highly redundant, even duplicated information, which is undesirable for users. Furthermore, the user's needs might be multi-faceted or ambiguous. The redundancy in top ranked results will reduce the chance to satisfy different users. 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.. In text summarization, candidate sentences of a summary are expected to be less redundant and cover different aspects of information delivered by the document. 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), subtopic diversity, cluster based centroids selecting, categorization based approach, and many other redundancy penalty approaches [34], [17], [30]. In this research 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 over data manifold, which can help find the most relevant and important data objects. We applied MRSP to two application tasks, update summarization and query recommendation. 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.

2. Manifold Rank

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. 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 . However, diversity is not considered in manifold ranking.

3. 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

4. Update Summarization

Update summarization is a temporal extension of topic focused multi document summarization, 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 in which some deep natural language processing techniques are
Leveraged to compress sentences another one is extractive summarization. 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.

5. Implementation

In this experiment, we construct the sentence manifold according to the pair-wise similarity values sim (xi, xj) between sentences xi and xj . xi is a term vector
Recording the *tf-isf* (term frequency-inverse
Sentence frequency) values of the sentence. The pair wise 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 Wij = sim(xi, xj) if there is an edge linking xi and xj . Let Wij = 0 to avoid self-loops in the graph.
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 dataset 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 dataset
A as an initial sink point, including summary of A, the most representative sentence of A, all the sentences from A, and an aggregated pseudo sentence 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 pseudo sentence from A as the representation of A.

5 . CONCLUSION

In this research paper, we propose 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.

6. REFERENCES

1.R.Agrawal and S.Gollapudi, “Diversifying Search Results,” Proc. Second ACM Int’l Conf. Web Search and Data Mining (WSDM ’09), pp. 5-14, 2009.

2] J. Allan 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.
[04] K. J arvelin and J. Kekalainen," Cumulated gain-based evaluation
of techniques", *ACM Trans. Inf. Syst.*, 20(4):422–446, 2002.
[05] K. Knight and D. Marcu," Statistics-based summarization – step one: Sentence compression", In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, pages 703–710.
AAAI Press / The MIT Press, 2000.
[06] Y. Lan and H. Li," Generalization analysis of list wise learning-to-rank algorithms", In *ICML '09: Proceedings*
of the 26th Annual International Conference on Machine Learning, pages 577–584, New York, NY, USA, 2009. ACM.
[07] L. Li and M. Kitsuregawa, "Query-URL bipartite based approach to personalized query recommendation", In
Proceedings of the 23rd national conference on Artificial intelligence, pages 1189–1194, 2008.