ISSN: 2320-1363

An efficient information retrieval by using Manifold Ranking with Sink Points

N.Vijayalakshmi*1, R.Pitchaiah*2

[1]M.Tech Scholar, Dept of CSE, Universal College of Engineering & Technology, Perecherla, Dist: Guntur, AP, India

[2]Associate Professor & HOD, Dept of CSE, Universal College of Engineering & Technology, Perecherla, Dist: Guntur, AP, India

ABSTRACT

Retrieving the relevant information for a particular searching is an important issue of information retrieval system. Ranking plays an important role in order to provide the relevant information according to the degree of relevance or importance. There is a chance of producing the redundant data. The traditional ranking algorithms don't concentrate on the diversity of information. We are proposing a new ranking algorithm called Manifold Ranking Sink Points to address the diversity problem and also produce the most relevant or important information. Our proposed algorithm uses the manifold ranking on the manifold data and provides the most relevant and important data over the data objects.

KEYWORDS: Information Retrieval, Diversity in ranking, manifold ranking sink points.

I.INTRODUCTION:

Retrieving the relevant information for a particular searching is an important issue of information retrieval system. Ranking is an important criterion to produce the relevant information from the bulk of data. Ranking is used for all the applications like information retrieval (IR), Data mining and Natural language processing. The basic working of ranking will sort the group of objects by using a simple function and displays the results on the order of relevance and importance. There are many traditional ranking algorithms which have a disadvantage of showing redundant data and

ISSN: 2320-1363

don't address diversity of the data objects. The redundant information will reduce to the user's satisfaction. Beyond the relevance and importance, diversity is also an important factor of searching and producing the results back. 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.

II.RELATED WORK

There are many approaches which uses ranking algorithms for information retrieval which produces the relevant and important data but suffering from the data diversity problem. The diversity problem was studied by many researchers. They produce different approaches to address the problem of diversity such as maximum marginal relevance (MMR)[1], subtopic diversity[2], cluster based cancroids selecting[3], categorization based approach[4], and many other redundancy penalty approaches [5],[6], [7]. The drawback with these methods is the separation of relevance and diversity. i.e these approaches unable to combine the two at one place.

We propose a new algorithm called manifold ranking with sink points (MRSP) which address the diversity problem and provides the results which are relevant data which improves the user satisfaction. This approach uses a manifold ranking process 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

ISSN: 2320-1363

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 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.

III.PROPOSED METHODOLOGY

3.1 Manifold Ranking with Sink Points:

MRSP assumes all the data and query objects are points sampled from a low-dimensional manifold and leverages a manifold ranking process 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. MRSP is a an iterative algorithm as explained as first find more relevant points add them to sink point and update their scores.

3.2 Algorithm:

Step1: Initialize the set of sink points X_s as empty.

Step2: Form the affinity matrix W for the data manifold, where Wij = sim(xi, xj) if there is an edge linking xi and xj . Note that sim(xi, xj) is the similarity between objects xi and xj .

Step3: 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 i-th row of W.

Step4: Repeat the following steps if |Xs| < 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 xi ∈ Xs and 1 otherwise.

ISSN: 2320-1363

(b) Let f_i^* denote the limit of the sequence $\{f_i(t)\}$. Rank points xi ∈ Xr 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 Xr to Xs

Step5: Return the sink points in the order that they were selected into Xs from Xr.

IV.RESULTS AND DISCUSSIONS

We apply our MRSP algorithm to a couple of real applications: update summarization and query recommendation. 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.

V.CONCLUSION

Our proposed manifold ranking with sink points (MRSP) will improves the search results. The proposed algorithm produces the most relevant and important information and also addresses the diversity problem. The search results improve the user's satisfaction. Finally we had applied our approach for Update summarization and query recommendation.

VI.REFERENCES

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

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

ISSN: 2320-1363

1. D. R. Radev, H. Jing, and M. Budzikowska. Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. In *NAACL-ANLP 2000 Workshop on Automatic summarization*, pages 21–30, Morristown, NJ, USA, 2000.
2. R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong. Diversifying search results. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, WSDM '09, pages 5–14, New York, NY, USA, 2009. ACM.
3. 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. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 504–511, New York, NY, USA, 2005. ACM.
4. W. Li, F. Wei, Q. Lu, and Y. He. PNR2: Ranking sentences with positive and negative reinforcement for query-oriented update summarization. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 489–496, Manchester, UK, August 2008.
5. 7 X. Wan, J. Yang, and J. Xiao. Manifold-ranking based topic focused multi-document summarization. In *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 2903–2908, Hyderabad, India, January 6-12 2007.
6. R. L. Santos, C. Macdonald, and I. Ounis. Exploiting query reformulations for web search result diversification. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 881–890, New York, NY, USA, 2010. ACM.
7. D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Scholkopf. Learning with local and global consistency. In S. Thrun,L. Saul, and B. Sch¨olkopf, editors, *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.
8. D. Zhou, J.Weston, A. Gretton, O. Bousquet, and B. Scholkopf. Ranking on data manifolds. In S. Thrun, L. Saul, and B. Sch¨olkopf, editors, *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.