



ADVANCEMENT TOWARDS DIVERSIFICATION OF SEARCH RESULTS

V.Veerabrahma Chari¹, T.S.Suhasini²

¹M.Tech Student, Dept of CSE, CMR Institute of Technology, Kandlakoya Medchal, Hyderabad, India

²Assistant Professor, Dept of CSE, CMR Institute of Technology, Kandlakoya Medchal, Hyderabad, India

ABSTRACT:

Summarization is the most effectual ways to combine information to keep away from information overload. There is a huge body of research in making structured representation in support of news event like timeline summarization. Query recommendations are significant tool that assist search engines users in information looking for activities, moreover known as search mission. Query recommendation is a mainstay mission for huge industrial search engines. Most of exertion on query recommendation spotlight on method of query resemblance, where query log data is extensively used. Ranking has plentiful applications in information recovery, and natural language processing. Manifold Ranking with Sink Points that was introduced to tackle diversity in addition to relevance in a combined system. The system makes uses a manifold ranking procedure over data manifold, which assist in searching most pertinent and significant data objects.

Keywords: *manifold ranking with sink points, query recommendation.*

1. INTRODUCTION:

Ranking on data manifolds are projected by in which data objects are supposed to be points modelled from low-dimensional manifold entrenched in a high-dimensional

euclidean space [8]. Manifold ranking as shown in fig1 provide elevated ranks to nodes that are secure to queries on manifold. Relevance in addition to significance is well reasonable in manifold ranking, comparable to Personalized PageRank [2]. Diversity is

not measured in manifold ranking. Ranking has plentiful applications in information recovery, and natural language processing [15]. In numerous real situations, the ranking difficulty is definite as: specified a collection of data objects, a ranking model sorts objects in grouping consistent with their degrees of relevance, or preferences [5] [14]. The redundancy in top ranked results will decrease possibility to convince dissimilar users. Beyond relevance as well as significance, diversity is acknowledged as a critical standard in ranking. In text summarization, candidate sentence of summary are likely to be less outmoded and envelop dissimilar aspects of information distributed by document [9]. In e-commerce, a listing of applicable but typical products is functional for user's to glance through and construct a purchase. Top ranked consequence is likely to get across as slight outmoded information as promising, and envelop as many feature as feasible [13]. Numerous real application responsibilities insist assortment in ranking. In query recommendation, the suggested query has to detain dissimilar query intents of dissimilar users [6]. Query recommendation intends to make available unconventional queries to assist users look for and also get better

usability of search engines. It has been engaged like a core efficacy by numerous industrial search engines [12]. Most of exertion on query recommendation spotlight on method of query resemblance, where query log data is extensively used. The manifold ranking algorithm is projected basis on assumptions such as close by data are expected to include close ranking score; as well as data on similar structure are probable to encompass secure ranking scores [14].

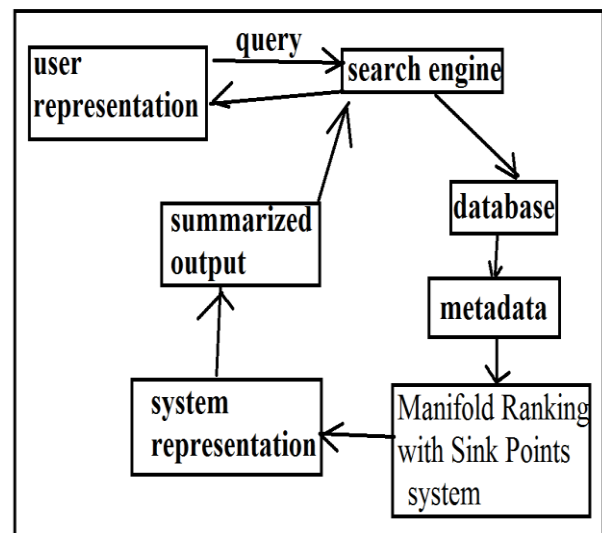


Fig1: An overview of system of Manifold Ranking with Sink Points.

II. LITERATURE SURVEY:

1. **Xue-Qi Cheng, Pan Du** [1] suggests a novel advance, named Manifold Ranking with Sink Points that was introduced to tackle diversity in addition to relevance in a

combined system. The system makes uses a manifold ranking procedure over data manifold, which assist in searching most pertinent and significant data objects. We commence into manifold sink points, which are objects whose scores of ranking are unchanging at lowest score throughout the procedure of manifold ranking. Ranking scores of additional objects secure to sink points will be obviously penalized throughout the ranking procedure based on essential manifold. By turning objects of ranked into sink points within data manifold, we can efficiently prevent outmoded objects from acceptance of a high rank. We can detain diversity in addition to significance throughout the ranking procedure. Introduced system has not merely a nice convergence property, however also a fulfilling optimization elucidation. We applied Manifold Ranking with Sink Points to two tasks of application such as update summarization as well as query proposal. Update summarization intend to review advanced information contained in novel document set specified a precedent document set. The mission of query proposal is to make available different queries to assist user's exploration and

moreover get better usability of search engines.

2. Giang Binh Tran [4] put forward that summarization is the most effectual ways to combine information to keep away from information overload. It is moreover expensive for numerous real life tasks, for instance maintaining information in support of future generation. Document hierarchical classification, News document hierarchical classification has been expansively considered for long time. The frequent approaches categorize texts into a hierarchical considered category predefined through human or inferred from knowledge ontology.

Machine learning algorithms such as SVM, NaiveBayes are extensively used. This research is a suggestion to take out the hierarchical relation among news events. There is a huge body of research in making structured representation in support of news event like timeline summarization. The system of timeline generation aims on extracting significant events and position them on a chronological time extent. Text summarization is expansively investigated and there are several methods to organize this difficulty. We plan to make available

more controlled representation which can detain important semantic features concerning the article collection.

3. P. Boldi, F. Bonchi [10] suggests that query recommendations are significant tool that assist search engines users in information looking for activities, moreover known as search mission. Recommendations are naturally queries comparable to innovative one, and they are typically obtained by means of analyzing query logs, for instance discovery of recommendations by means of clustering concerning queries or by means of recognizing recurrent re-phrasings. The most important source of information in support of building search supporting systems are query logs which are extremely large, incredibly diverse heavy-tailed as well as noisy. Distilling behavioural prototype from query logs is a significant step towards getting better the service offered by search engines and on the way to developing pioneering web-search paradigms. Query recommendation is a mainstay mission for huge industrial search engines. Generally effort on query recommendation is focused on process of query resemblance that can be employed for query extension or else query clustering.

One present research line we extend attempts to understand the unknown semantics of user connections by search engines by means of extracting information from query log in steps.

4. R. Agrawal, S. Gollapudi [3] recommends that the consequence of result diversification has been predictable since early effort on information recovery. The fundamental principle is that significance of set of documents depends not merely on individual relevance of its associates, but also on relation to one another. The document set has to appropriately account for welfare of general user population. The majority of current exertion on result diversification makes at finest merely a tacit use of topics of documents, and diversification happens by means of resemblance functions over documents. We spotlight on how to vary search results building unambiguous use of information concerning the topics query or else the documents might refer to. Up to date user learning suggests that efforts of individual users are typically focussed in the direction of finding additional information on precise topics of attention, rather than an undirected mission for any novel information. We show

the common problem is NP-hard, however the objective utility admits a sub modularity construction that can be exploited for functioning of a superior approximation algorithm. The algorithm is best possible in usual special case where every document belong to accurately one group. As classical information retrieval metrics for instance normalized discounted cumulative gain, mean average precision, as well as mean reciprocal rank do not take into report the significance of diversification, we put forward a generalization of metrics to assess the search results.

5. J. Allan, R. Gupta [7] suggests that there has been extremely diminutive work on time-based summarization up to now. Several researchers have spotlight on how to remove temporal expressions from text, in search of and regularize references to elapsed times. That exertion is significant for analyzing material of text however does not unswervingly tackle summarization itself. Document summaries are hard to assess, since for the most part of applications there are frequent summary that are of uniformly high quality. Summarization method influences a extensive range of Natural Language

Processing as well as discourse information. Some spotlight first and foremost on method that has been expanded in Information recovery although the majority attempt to influence IR approaches and some features of NLP. Several summarization exertions were focused on events. Maybury's effort on event data is dissimilar than this effort since he was spotlighted on events from application data to a certain extent than on events in news topics.

III. CONCLUSION:

There has been extremely diminutive work on time-based summarization up to now. Several researchers have spotlight on how to remove temporal expressions from text, in search of and regularize references to elapsed times. Several summarization exertions were focused on events. The mission of query proposal is to make available different queries to assist user's exploration and moreover get better usability of search engines. Generally effort on query recommendation is focused on process of query resemblance that can be employed for query extension or else query clustering. The majority of current exertion on result diversification makes at finest merely a tacit use of topics of documents,

and diversification happens by means of resemblance functions over documents.

REFERENCES:

- [1] "Ranking on Data Manifold with Sink Points", Xue-Qi Cheng, Pan Du, Jiafeng Guo, Xiaofei Zhu, and Yixin Chen, 2013
- [2] 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.
- [3] R. Agrawal, S. Gollapudi, A. Halverson, and S. Jeong, "Diversifying Search Results," Proc. Second ACM Int'l Conf. Web Search and Data Mining (WSDM '09), pp. 5-14, 2009.
- [4] Structured Summarization for News Events Giang Binh Tran, 2013
- [5] C.L. Clarke, M. Kolla, G.V. Cormack, O. Vechtomova, A. Ashkan, S. Bu"tcher, 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Scho"lkopf, "Ranking on Data Manifolds," Advances in Neural Information Processing Systems 16. S. Thrun, L. Saul, and B. Scho"lkopf, eds., MIT Press, 2004
- [15] 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.