Study of Mrsp – A Comparison Study of Diversity, Relevance, Importance in Ranking

Neerukonda Venkatesh  
PG scholar  
Department of IT  
Geethanjali college of Engg. & Technology  
Hyderabad, India

B.Srinivas  
Associate Professor  
Department of IT  
Geethanjali college of Engg. & Technology  
Hyderabad, India

Abstract: Information retrieval is an everyday task for different domain area users for specific task related to their area of interest for knowledge discovery (KDD), Decision makings etc. Data mining techniques plays a vital role for extracting information related to user search. Search Engines involves in retrieving of Users’ queries by searching in multiple databases. As the users of the technology growing rapidly, the enormous collection of data in sort of data links & maintaining of that data by storing in databases became a big problem for companies to retrieve the Information Relevance to the given query, users are totally depending on the retrieval data. Here comes the actual problem in estimating the Relevance, Importance and divergence of the result obtained. Ranking is the primary concept in displaying any result on Web based Systems (WS), Automated Systems (AS), Information Retrieval Systems (IRS) by fulfilling the common properties. In this paper we discussed in ranking properties, divergence problems and studied an approach -Manifold ranking with sink points (MRSP) to address diversity as well as relevance and importance in ranking.

Keywords: KDD, IRS, Ranking, MRSP.

I. INTRODUCTION

Ranking is typically used for displaying the Webpages at client-side. Ranking has many applications in real world scenario in various fields like Data mining, Information Retrieval and Natural Language Processing. In order to retrieve a web page related to user query, which page has to be retrieved and displayed for user is the process at search engine done based on Relevance, Importance of given query. Redundancy, Diversity in results noticed as the basic problems of Ranking. Researchers from various domains have proposed many approaches to address this problem, such as Maximum Marginal Relevance (MMR) [1], subtopic diversity [2], [4], cluster-based centroids selecting [3], categorization-based approach [1], and many other redundancy penalty approaches [6], [5], [7]. However, these methods often treat relevance and diversity separately in the ranking algorithm, sometimes with additional heuristic procedures. In this paper, we studied 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. Meanwhile, we introduce into the manifold sink points, which are objects whose ranking scores are fixed at the minimum score 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

II. RANKING ON DATA MANIFOLDS

Manifold Ranking algorithm works based on two conditions:  
1) Nearby data are likely to have close ranking scores  
2) Data on the same structure are likely to have close ranking scores.

1) Description of Ranking Algorithm

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

Fig 1: Indication of page Rankings flow.
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.

III. GOALS OF 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.

Goals of update summarization:

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.

REFERENCES


