International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

# Ranking Models and Learning to Rank: A Survey

**Snehal Ukarande[1], Ashish Manwatkar[2]**

[1 2]Indira College of Engineering and Management, Pune, Maharashtra, India

**Abstract:** ***Current era is of fastest information retrieval. There are lots of research methodologies which are arising to give the fastest and correct result set. This paper mainly focuses on survey of the ranking models and learning to rank technique for giving the effective and efficient information retrieval. Learning to rank is flattering progressively more trendy research area in machine learning. The problem of ranking aims to encourage an ordering or inclination of relation among a set of instances in the input space. Learning to rank for information retrieval has gained a lot of interest in the recent years as ranking is the central problem in many information retrieval applications, like document retrieval, multimedia retrieval, text summarization, collaborative filtering, question answering and online advertising machine translation etc. The tremendously large size of the Web documents makes it usually impracticable for the common users for finding their desired information by surfing the net. As a result, effective and efficient information retrieval is being more important and also search engine (information retrieval system) has turned out to be a vital tool for people to locate their needed information.***

**Keywords:** Ranking Model, Learning to Rank, Information Retrieval, Data Mining.

## 1. Introduction

Learning to rank is a quite new research area which has evolved in the past decade. The search engines are necessary for finding and exploring information on the Web and other information systems. To a better extent the ranking function determines the excellence of search engines, used to create the results according to user's query as ranking is the heart of information retrieval system. When user queries, the documents have to be ranked according to the relevance to the query. Various machine learning algorithms are being used to learn the ranking function. Thus, Ranking has widespread applications such as commercial search engines and recommendation system that can find out relevance between the relevant documents in context of given user's query and place them in order of their significance in rank list. Such classification explores the queries and documents are given where each query is associated with a Perfect ranking list of the documents. The model for ranking is then fashioned using the classification process according to the given query.

In distinction, learning to rank methods in information retrieval permits retrieval systems to incorporate hundreds or even thousands of arbitrarily defined features. Importantly, these approaches, without human intervention, learn the most effective combination of the features in the ranking function depending on the available data for classification. The evaluation metrics are required to compute the quality of search engine that is one of the most normally used metric in web search ranking i.e Discounted Cumulative Gain (DCG), it is used to compute ranking quality of search engines. The information retrieval is often used to evaluate usefulness of web search engine algorithms or other related applications. DCG measures the usefulness, of document based on its perfect position in the rank list. If the relevant document is in lower location then it is not more useful for the user to gain knowledge. The principle is to integrate both query level selection and document level selection for ranking and introduce an expected discounted cumulative gain (DCG) loss optimization (ELO-DCG) algorithm, for selecting most informative and relevant document associated to query.

## 2. Learning to Rank Survey

Learning to rank [2] has three widespread approaches they are: Point wise approaches, Pair wise approaches, and List wise approaches. These three different approaches can be taught to rank in different ways. With the intention of ranking different input and output spaces may be defined, different hypotheses may be used, and employs different loss functions. The Point wise approaches are the earlier approaches [2] the basic hypotheses of this approach is used for mapping the document's ordinal scales into numeric values using regression and classification method, it try to compare the relevance result of every two documents, then comparison result is produced. Based on that result the document will be ranked. Binary classifier method is used by pair wise approach that will tell which document is better in a given pair of documents.

Goal of using binary classifier is minimizing average number of inversions for ranking functions. The List wise approaches [3] is alike with the basic idea of pair wise approach, it straightforwardly compare the relevance list of documents based on query, as an alternative of trying to get ranking score for each document independently. It uses Ad a Rank and soft Rank algorithms for giving rank. Compared with traditional active learning algorithm; there is still incomplete work in the active learning for ranking in recent years. The problem of text selection based on query in ranking is studied by Donmez and Carbonell [3]. The ambiguity sampling is simple and common strategy in active learning, the issue in sampling is that the algorithm is selecting queries for which the label uncertainty samples have highest relevance score [4].The main disadvantage in this type of approach is noise as well as of variance. Active learning algorithm minimizes the noise and reduces variance proposed in [5]. Query by Committee algorithm [8] is using noise free classification function. Another common approach for active learning is to choose query that once added to training set leads to large increase in the objective function value that is being optimized [6]. Most of the other ranking algorithms such as Rank SVM [9] and Rank Boost [7] suggests to add the most relevant pairs of documents to the training set, the document's predicted relevance scores are

very close under the current ranking models. In the terms of binary relevance, greedy algorithm [1] is proposed which selects the document which differentiates two different ranking systems in terms of average precision. The comparison of relevant document selection methodologies in learning to rank are found in [8]. L. Yang, L. Wang [4] proposed greedy query selection algorithm which minimizes query density and query diversity. Some empirical and theoretical work related to query sampling are found in[5] the results shows that better having more queries but less number of documents per query than having more documents and less queries.

## 3. Ranking Models Survey

Most of the internet users rely on search engines for extracting information by providing a query from any walk of life. The search engines processes these queries and a certain information retrieval or mining algorithm is applied to obtain the cluster of documents relevant to the query. After the retrieval of these documents, an important task is to present these documents in the list where documents at the top are the ones considered more relevant for the user. This task of collecting most relevant document at top of the list is called ranking of documents.

There are basically two types of ranking models: static ranking models and dynamic ranking models. In earlier days ranking algorithms were based on prior information about the websites. PageRank, SALSA,HITS, RankNet and fRank are examples of static algorithms. These use static features of web pages and hence are termed here as Static Ranking algorithms. The static ranking algorithms don't take into consideration of the interaction with user and faces issues like query ambiguity and diversity in intent of user. There is an inherent trade-off among number of results provided for user intent and number of intents retrieved .A way to combine otherwise contradictory goals of result diversification and high recall is provided by dynamic ranking. These algorithms runs the interaction with the user to know his intent amongst the various possible intents, or they try to reorder the results of first retrieval process and provides refined results to the user. They focus on both the relevance and diversity

Ranking can be applied at different applications for eg. Huang-Chia Shih, Jenq-Neng Hwang and Chung-Lin Huang has proposed a system which uses content based attention ranking Using Visual and Contextual Attention Model for Baseball Videos[11]. They have analyzed how people are excited about the watched video content and proposes a content-driven attention ranking strategy which is enablling client users to iteratively browse the video according to their preference. The attention rank (AR) algorithm, which is extended from the Google PageRank algorithm that sorts the websites based on the importance, can effeciently measure the user interest (UI) level for each video frame. Integration of the object-based visual attention model (VAM) with context attention model (CAM) derives the degree of attention, which can more reliably takes the advantage of the human perception characteristics, as well as can effectively identify which video contents can attract users' attention. The information of users' feedback is utilized in re-ranking procedure to further improves the retrieving accuracy. The proposed algorithm is specifically evaluated on broadcasted baseball videos [11].

Ranking models can be applied to domain specific search[12]. With the explosive emergence of vertical search domains, application of the the broad-based ranking model directly to different domains is no longer desirable due to domain differences, during building of a unique ranking model for each domain is both laborious for labeling data and time consuming for training models. Bo Geng in his paper, addressed these difficulties by proposing a regularization-based algorithm called ranking adaptating RA-SVM, through which we can adapt an existing ranking model to a new domain, so that the quantity of labeled data and the training cost is reduced during the performance is still guaranteed. This algorithm requires the prediction from the existing ranking models, rather than their internal representation or the data from auxiliary domains. In addition, authors assume that documents having similarity in the domain-specific feature space should have consistent rankings, and add some constraints for controlling the margin and slack variables of RA-SVM adaptively. Finally, ranking adaptability measurement is proposed to quantitatively estimating if an existing ranking model can be adapted to a new domain[12].

## 4. Architecture of document retrieval system

The documents present in database will be searched by the user. According to the need and query, the previously ranked documents will be displayed as result to the user. If the documents are properly ranked then the loss of data is optimized while searching relevant data. The user sometimes searches for a document and due to improper ranking the most relevant data is also not able to be retrieved. Hence proper ranking models should be used in order to retrieve total relevant data thereby reducing the loss.

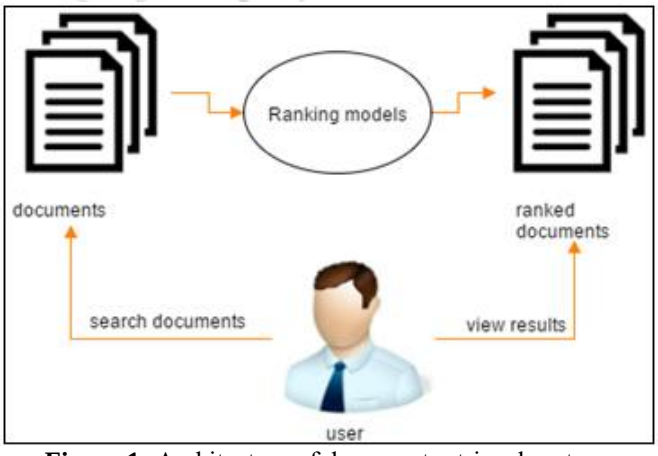


**Figure 1:** Architecture of document retrieval system

## 5. Conclusions

As technology improves each day new developments are continuously infiltrating our lives. Research in learning to rank is a friendly process and the must of ranking change every day depending on the requirements from the user. Active learning for ranking differs from Active learning for

classification and regression including learning for ranking that has some unique features. There are many ranking algorithm which are all time consuming and also cost much in obtaining labeled data compared with those algorithm Expected loss optimization for query and document level ranking by active learning performs efficiently by providing the user the most informative documents for their references [10].

## References

[1] Bo Long, Jiang Bian Olivier Chapelle, Ya Zhang, Yoshiyuki Inagaki, and Yi Chang, "Active Learning for Ranking through Expected Loss Optimization",IEEE transactions on knowledge and data engineering, VOL. 27, NO. 5, MAY 2015.
[2] B. Qian, H. Li, J. Wang, X. Wang, and I. Davidson, "Active Learning to Rank using Pairwise Supervision," In Proc. 13th SIAM Int.Conf. Data Mining, 2013, pp. 297–305.
[3] P. Donmez and J. G. Carbonell,"Optimizing estimated loss reduction for active sampling in rank learning", In ICML '08: Proceedings of the 25th international conference on Machine learning, pages 248- 255, New York, NY, USA, 2008. ACM.
[4] D. Lewis and W. Gale, "Training text classifiers by uncertainty sampling", In Proceedings of the 17thAnnual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 3-12, 1994.
[5] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models", In G. Tesauro, D. Touretzky, and T. Leen, editors, Advances in Neural Information Processing Systems, volume 7,pages 705-712. the MIT Press, 1995.
[6] C. Campbell, N. Cristianini, and A. Smola, "Query learning with large margin classifiers", In Proceedings of the Seventeenth International Conference on Machine Learning, pages 111-118. Morgan Kaufmann, 2000.
[7] Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer,"An efficient boosting algorithm for combining preferences", Journal of Machine Learning Research, 4:933-969,2003.
[8] Y. Freund, H. S. Seung, E. Shamir, and N, "Tishby Selective sampling using the query by committee algorithm",Machine Learning, 28(2- 3):133-168, 1997
[9] A. Aslam, E. Kanoulas, V. Pavlu, S. Savev, and E. Yilmaz, "Document selection methodologies for efficient and effective learning-to-rank," In Proc. 32nd Int. ACM SIGIR Conf. Res. Develop.Inform. Retrieval, 2009, pp. 468–475.
[10] Aditi Sharma, Nishtha Adhav, Anju Mishra ,"A Survey : Static and Dynamic Ranking " , International Journal of Computer Applications (0975 -8887) Volume 70 No-14 ,May 2013.
[11] Huang-Chia Shih, Jenq-Neng Hwang, Fellow and Chung-Lin Huang, "content based attention ranking Using Visual and Contextual Attention Model for Baseball Videos", IEEE transactions on multimedia, vol. 11, NO. 2, feb 2009.
[12] Bo Geng, Linjun Yang, Chao Xu, and Xian-Sheng Hua, "Ranking Model Adaptation for Domain-Specific Search", IEEE transactions on knowledge and data engineering,vol. 24, no. 4, april 2012.