



# “Active Learning for Ranking through Expected Loss Optimization” based on Data Mining

Parkhi Aishwarya<sup>1</sup>, Patil Chaitanya<sup>2</sup>, Shinde Madhuri<sup>3</sup>

Department of Computer Engineering Pune, India<sup>1,2,3</sup>

**Abstract:** Learning to rank arises in many data mining applications, ranging from web search engine, online advertising to recommendation system. In learning to rank, the performance of a ranking model is strongly affected by the number of labelled examples in the training set; on the other hand, obtaining labelled examples for training data is very expensive and time consuming. This presents a great need for the active learning approaches to select most informative examples for ranking learning; however, in the literature there is still very limited work to address active learning for ranking. In this paper, we propose a general active learning framework, expected loss optimization (ELO), for ranking. The ELO framework is applicable to a wide range of ranking functions. Under this framework, we derive a novel algorithm, expected discounted cumulative gain (DCG) loss optimization (ELO-DCG), to select most informative examples. Then, we investigate both query and document level active learning for ranking and propose a twostage ELO-DCG algorithm which incorporate both query and document selection into active learning. Furthermore, we show that it is flexible for the algorithm to deal with the skewed grade distribution problem with the modification of the loss function. Extensive experiments on realworld web search data sets have demonstrated great potential and effectiveness of the proposed framework and algorithms.

**Keywords:** Active Learning, Data Mining.

## I. INTRODUCTION

Ranking is the core component of many important information retrieval problems, such as web search, recommendation computational advertising. Learning to rank represents an important class of supervised machine learning tasks with the goal of automatically constructing ranking functions from training data. As many other supervised machine learning problems, the quality of a ranking function is highly correlated with the amount of labelled data used to train the function. Due to the complexity of many ranking problems, a large amount of labelled training examples is usually required to learn a high quality ranking function. However, in most applications, while it is easy to collect unlabelled samples, it is very expensive and time consuming to label the samples. Existing algorithms for learning to rank may be categorized into three groups point wise approach [8], pairwise approach [26], and list wise approach [22].

## II. RELATED WORK

The main motivation for active learning is that it usually requires time and/or money for the human expert to label examples and those resources should not be wasted to label oninformative samples, but be spent on interesting ones. Optimal experimental design [12] is closely related to active learning as it attempts to find a set of points such that the variance of the estimate is minimized. In contrast to this “batch” formulation, the term active learning often refers to an incremental strategy [7]. There has been various types of strategies for active learning that we now review. A comprehensive survey can be found in [21]. The simplest and maybe most common strategy is uncertainty sampling [18], where the active learning algorithm queries points for which the label uncertainty is the highest. The drawback of this type of approach is that it often mixes two types of uncertainties, the one stemming from the noise and the variance. The noise is something intrinsic to the learning problem which does not depend on the size of the training set. An active learner should not spend too much effort in querying points in noisy regions of the input space. On the other hand, the variance is the uncertainty in the model parameters resulting from the finiteness of the training set. Active learning should thus try to minimize this variance and this was first proposed in [7].

### A. EXPECTED LOSS OPTIMIZATION FOR ACTIVE LEARNING

As explained in the previous section, a natural strategy for active learning is based on variance minimization. The variance, in the context of regression, stems from the uncertainty in the prediction due to the finiteness of the training set. Cohen et al. [7] proposes to select the next instance to be labelled as the one with the highest variance. However, this approach applies only to regression and we aim at generalizing it through the Bayesian expected loss [3]. In the rest of the section, we first review Bayesian decision theory in Section 3.1 and then introduce the expected loss



optimization principle for active learning. In Section 3.2 we show that in the cases of classification and regression, applying ELO turns out to be equivalent to standard active learning method. Finally, we present ELO for ranking in Section 3.3.

### B. ELO for Ranking

In the case of ranking, the input instance is a query and a set of documents associated with it, while the output is a vector of relevance scores. If the query  $q$  has  $n$  documents, let us denote by  $X_q := (x_1; \dots; x_n)$  the feature vectors describing these (query, document) pairs and by  $Y := (y_1; \dots; y_n)$  their labels. As before we have a predictive distribution  $p(Y|X_q, D)$ . Unlike active learning for classification and regression, active learning for ranking can select examples at different levels. One is query level, which selects a query with all associated documents  $X_q$ ; the other one is document level, which selects documents  $X_i$  individually.

### C. Two-Stage Active Learning

Both query level and document level active learning have their own drawbacks. Since query level active learning selects all documents associated with a query, it tends to include non-informative documents when there are a large number of documents associated with each query. For example, in Websearch applications, there are large amount of Web documents associated for a query; most of them are non-informative, since the quality of a ranking function is mainly measured by its ranking output on a small number of top ranked Web documents. On the other hand, document level active learning selects documents individually. This selection process implies unrealistic assumption that documents are independent, which leads to some undesirable results. For example, an informative query could be missed if none of its documents is selected; or only one document is selected for a query, which is not a good example in ranking learning.

## III. PROPOSED SYSTEM

We first propose a general active learning framework, expected loss optimization (ELO), and apply it to ranking. The proposed algorithm is further extended to a two-stage active learning schema to seamlessly integrate query level and document level data selection.

## IV. LITERATURE SURVEY

**Authors-** Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer.

AdaRank: A Boosting Algorithm for Information Retrieval

AdaRank develops a new learning algorithm that can directly optimize any performance measure used in document retrieval. In document retrieval, usually ranking results are evaluated in terms of performance measures such as MAP (Mean Average Precision) and NDCG (Normalized Discounted Cumulative Gain). AdaRank algorithm can iteratively optimize an exponential loss function based on any of IR performance measures. AdaRank can be viewed as a machine learning method for ranking model tuning [5].

**Authors-** Sarthak Jain, Anant Vaibhav, Lovely Goyal

SVM Selective Sampling for Ranking with Application to Data Retrieval

It produced practical applications in information retrieval. SVM selective sampling technique is used for learning ranking function. Selective sampling technique is to select the most ambiguous samples for ranking at each round, so that the users feedback on those samples will maximize the degree of learning. In this sampling technique significantly reduces the labelling effort to learn an accurate SVM ranking function and it apply method to the data retrieval application [2].

**Authors-** Monika M Patel a, Mehul A Jajal, Dixita B vataliya

Document Selection Methodologies for Efficient and Effective Learning-to-Rank

Document selection methodology employ a number of document selection methodologies, widely used in the context of evaluation depth-k pooling, sampling and active learning. Document selection showed that first the proportion of relevant documents to non-relevant documents and second the similarity between relevant and non-relevant documents in the data sets [4].

**Authors-** Pavithra.D, Ranjith Balakrishnan

Active Learning for Ranking through Expected loss optimization

Active learning for ranking through expected loss optimization select most informative examples for ranking learning. In ranking represents an important class of supervised machine learning tasks with the goal of automatically constructing ranking functions from training data [1].



**Authors-** Vladimir Vujovic, Mirjana Maksimovic

Optimizing Search Engines using Click through Data

Optimizing search engine presents an approach to learning retrieval functions by analysing which links the users click on in the presented ranking. In this define what click through data is, how it can be recorded and how it can be used to generate training examples in the form of preferences? Click through data can provide training data in the form of relative preferences. Based on a new formulation of the learning problem in information retrieval, this derives an algorithm for learning ranking function [7].

## V. SYSTEM ARCHITECTURE

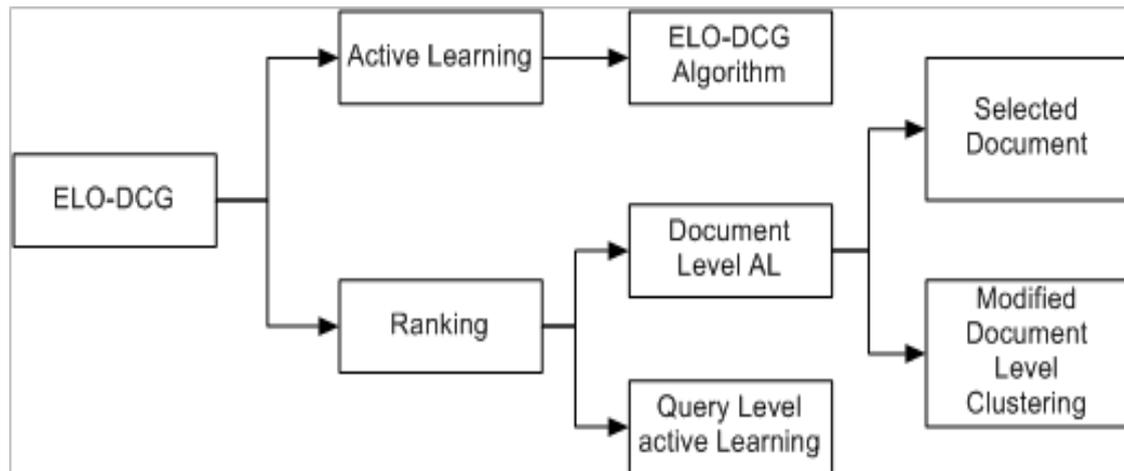


Fig. System Architecture

## VI. CONCLUSION

We propose a general expected loss optimization framework for ranking, which is applicable to active learning scenarios for various ranking learners. Under ELO framework, we derive novel algorithms, query level ELO-DCG and document level ELO-DCG, to select most informative examples to minimize the expected DCG loss. We propose a two stage active learning algorithm to select the most effective examples for the most effective queries. We further extend the proposed algorithm to deal with the typical skew grade distribution problem in learning to rank. Extensive experiments on real-world web search data sets have demonstrated great potential and effectiveness of the proposed framework and algorithms.

## REFERENCES

- [1] J. Guiver and E. Snelson, "Learning to rank with Soft Rank and Gaussian processes," in Proc. 31st Annu. Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval, 2008, pp. 259–266.
- [2] E. Yilmaz and S. Robertson, "Deep versus shallow judgments in learning to rank," in Proc. 32nd Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval, 2009, pp. 662–663.
- [3] Donmez and J. G. Carbonell, "Optimizing estimated loss reduction for active sampling in rank learning," in Proc. 25th Int. Conf. Mach. Learn., 2008, pp. 248–255.
- [4] 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.
- [5] Cortes, M. Mohri, and A. Rastogi. Magnitude-preserving ranking algorithms. In Proceedings of the 24th ICML, 2007.
- [6] Z. Zheng, K. Chen, G. Sun, and H. Zha. A regression framework for learning ranking functions using relative relevance judgments. In Proceedings of the 30th ACM SIGIR conference, 2007.
- [7] 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.
- [8] . Dagan and S. P. Engelson. Committee-based sampling for training probabilistic classifiers. In In Proceedings of the Twelfth International Conference on Machine Learning, pages 150–157. Morgan Kaufmann, 1995.